Adaptive processing of thin structures to augment segmentation of dual-channel structural MRI of the human brain

James Withers

Doctor of Philosophy

Institute for Adaptive and Neural Computation

School of Informatics

University of Edinburgh

2010

Abstract

This thesis presents a method for the segmentation of dual-channel structural magnetic resonance imaging (MRI) volumes of the human brain into four tissue classes. The state-of-the-art FSL FAST segmentation software (Zhang *et al.*, 2001) is in widespread clinical use, and so it is considered a benchmark. A significant proportion of FAST's errors has been shown to be localised to cortical sulci and blood vessels; this issue has driven the developments in this thesis, rather than any particular clinical demand.

The original theme lies in preserving and even restoring these thin structures, poorly resolved in typical clinical MRI. Bright plate-shaped sulci and dark tubular vessels are best contrasted from the other tissues using the T_2- and PD-weighted data, respectively. A contrasting tube detector algorithm (based on Frangi *et al.*, 1998) was adapted to detect both structures, with smoothing (based on Westin and Knutsson, 2006) of an intermediate tensor representation to ensure smoothness and fuller coverage of the maps.

The segmentation strategy required the MRI volumes to be upscaled to an artificial high resolution where a small partial volume label set would be valid and the segmentation process would be simplified. A resolution enhancement process (based on Salvado *et al.*, 2006) was significantly modified to smooth homogeneous regions and sharpen their boundaries in dual-channel data. In addition, it was able to preserve the mapped thin structures' intensities or restore them to pure tissue values. Finally, the segmentation phase employed a relaxation-based labelling optimisation process (based on Li *et al.*, 1997) to improve accuracy, rather than more efficient greedy methods which are typically used. The thin structure location prior maps and the resolution-enhanced data also helped improve the labelling accuracy, particularly around sulci and vessels.

Testing was performed on the aged LBC1936 clinical dataset and on younger brain volumes acquired at the SHEFC Brain Imaging Centre (Western General Hospital, Edinburgh, UK), as well as the BrainWeb phantom. Overall, the proposed methods rivalled and often improved segmentation accuracy compared to FAST, where the ground truth was produced by a radiologist using software designed for this project. The performance in pathological and atrophied brain volumes, and the differences with the original segmentation algorithm on which it was based (van Leemput *et al.*, 2003), were also examined. Among the suggestions for future development include a soft labelling consensus formation framework to mitigate rater bias in the ground truth, and contour-based models of the brain parenchyma to provide additional structural constraints.

Acknowledgements

I would like to thank the Neuroinformatics DTC at the University of Edinburgh for funding my masters course and PhD, as well as their main backers: the EPSRC, MRC and BBSRC. I would also like to express my gratitude to Dr Amos Storkey and Dr Mark Bastin for supervising me for the past four years.

The brain MRI data used in this thesis was collected at the SHEFC Brain Imaging Centre at the Western General Hospital (Edinburgh, UK). The radiology staff were very helpful and in particular I would like to thank Dr Andrew Farrall for contributing his time for the very tedious task of labelling over 25000 voxels individually.

I would like to acknowledge Dr Olivier Salvado for his correspondence and generous support for his initial work on artificially enhancing the resolution of MRI data, as well as for introducing me to differential geometry methods. Finally I would like to thank Dr Andrea Chai, without whom I would have given up on all this a long time ago!

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*James Withers*)

Table of Contents

List of Tables

List of Figures

Acronyms

Acronym	Definition
2D	2 Dimensional
3D	3 Dimensional
ADC	Apparent Diffusion Co-efficient
AIF	Arterial Input Function
ASL	Arterial Spin Labelling
BA	Brodmann's Area
BET	Brain Extraction Tool
BFC	Bias Field Correction
BG	Background/Blood
CFL	Courant-Friedrichs-Lewy condition
CMD	Constrained Mixel Dimension
CNR	Contrast-to-Noise Ratio
CSF	Cerebrospinal Fluid
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DTI	Diffusion Tensor Imaging
DTI	Diffusion Tensor Imaging
DWI	Diffusion-Weighted Imaging
E-step	Expectation step
EM	Expectation Maximisation
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
FAST	FSL Automated Segmentation Tool

Continued on next page

Continued from previous page

Acronym	Definition
FCM	Fuzzy *c*-Means
FLAIR	Fluid Attenuated Inversion Recovery
FLIRT	FSL Linear Image Registration Tool
fMRI	Functional Magnetic Resonance Imaging
FoM	Figure of Merit for edge overlap
FOV	Field Of View
FPR	False Positive Rate
FSL	FMRIB Software Library
FWHM	Full Width at Half Maximum
GM	Gray Matter
ICM	Iterated Conditional Modes
IR	Inversion Recovery
ITK	Insight Toolkit
LAPACK	Linear Algebra Package
LBC1936	Lothian Birth Cohort 1936
M-step	Maximisation step
MAE	Mean Absolute Error
MAP	Maximum A Posteriori
MEX	MATLAB Executable
MIPAV	Medical Image Processing, Analysis and Visualization
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NEX	Number of signal averages (or excitations)
NIfTI	Neuroimaging Informatics Technology Initiative
NPV	Negative Predictive Value
NSDR	Number of Site Disagreements Ratio
PD	Proton Density
PDE	Partial Differential Equation
PET	Positron Emission Tomography
PND	Principal Normal Direction

Continued on next page

Continued from previous page	
Acronym	**Definition**
PPV	Positive Predictive Value
PSF	Point Spread Function
PVE	Partial Volume Effect
R-step	Relaxation step
RAD	Reverse Anisotropic Diffusion
RAM	Random Access Memory
RD	Reverse Diffusion
RF	Radio Frequency
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SE	Spin Echo
SNR	signal-to-Noise Ratio
SPECT	Single Photon Emission Computed Tomography
SPM	Statistical Parametric Mapping
STAPLE	Simultaneous Truth And Performance Level Estimation
TE	Echo Time
TI	Inversion Time
TPR	True Positive Rate
TR	Repetition Time
VR space	Virchow-Robin perivascular space
VRE	Volume Resolution Enhancement
VTK	Visualization Toolkit
WM	White Matter

Chapter 1

Introduction

Non-invasive three-dimensional (3D) imaging is now widely used in clinical diagnosis, pathological investigation and treatment planning. Automated medical image analysis research seeks to reduce the labour costs, tedium, inconsistency and difficulty of visualisation associated with manual processing of the volumes. When an acceptable standard of accuracy (accordance with some ground truth) has been reached, automated tools tend to complete faster and with improved precision (reproducibility of the result) compared to manual expert processing, as argued in [Warfield et al., 2004].

This chapter will introduce a variety of background information on the topic of brain tissue segmentation using magnetic resonance imaging (MRI), which is necessary for a good understanding of both the goals and future chapters of this thesis. Particular mathematical concepts will be introduced in the later chapters as they are required. A primary goal of this project is to produce a clinically-useful tool; hence many descriptions and images in this chapter will address real-world concerns and data.

First, the acquisition of structural MRI of the brain will be described in a clinical context with particular attention paid to sources of segmentation confounds. A brief summary of MRI physics is presented to clarify issues caused by noise, artifacts and field strengths. Subsequently the appearance of brain anatomy in some common MRI weighting factors will be explored. A brief introduction to human brain tissue segmentation will follow, incorporating a break-down of some current issues in the field to be addressed in this project. Finally the chapter structure and the contributions of this thesis will be outlined alongside a list of publications and download locations of free, publically-available code.

1.1 Basics of Magnetic Resonance Imaging

In this section the basic terms of reference for MRI physics[1], the image artifacts encountered, and the tissue contrasts able to be achieved in clinical settings are described. Overall, the advantages of MRI over other imaging methods include:

- The range of acquisition trade-offs between improved image quality (in terms of contrast and noisiness), greater resolution, and shorter examinations;
- High spatial resolution possible in short examinations using current technology;
- The detection of temporal differences using rapid imaging sequences;
- Robustness to confounding factors;
- Excellent discrimination of soft tissue and anatomical structure, especially in the brain;
- Non-invasiveness for most types of acquisition.

1.1.1 Physics of acquisition

In the basic quantum model of MRI, nuclei interact with a magnetic field due to their intrinsic spin I. The most common nucleus studied is that of hydrogen (^{1}H) which is abundant in biological tissues. ^{1}H consists of a single proton with $I = \frac{1}{2}$ and angular momentum $p = \frac{h}{2\pi}I$, where h is Planck's constant. γ is the gyromagnetic ratio for ^{1}H which describes a magnetic dipole moment $\mu = \gamma p$. The combined μ of all nuclei in the sample is the bulk magnetisation M and the turning force experienced by nuclei in a field B is $L = MB$. The turning force can be visualised as precession about the main magnetic field B_0 (Fig 1.1A). Changes to M are typically visualised in an axis parallel to B_0, rotating in synchrony with the precessing magnetic moments: the rotating frame of reference (Fig 1.1B). This consists of three mutually orthogonal spatial axes, where the Z axis (pointing up) is in the same direction as B_0.

μ can initially be parallel or anti-parallel to B_0, corresponding to low or high energy states respectively. The energy difference between the states is $\Delta E = \frac{L}{I}$, and transitions between them can be induced by applying a pulse with this energy at the Larmor frequency $\upsilon_L = \frac{\omega_L}{2\pi}$. $\omega_L = \gamma B_0$ is the angular resonant frequency which forms a major part

[1] For a more complete introduction to MRI, the reader is referred to [McRobbie et al., 2007].

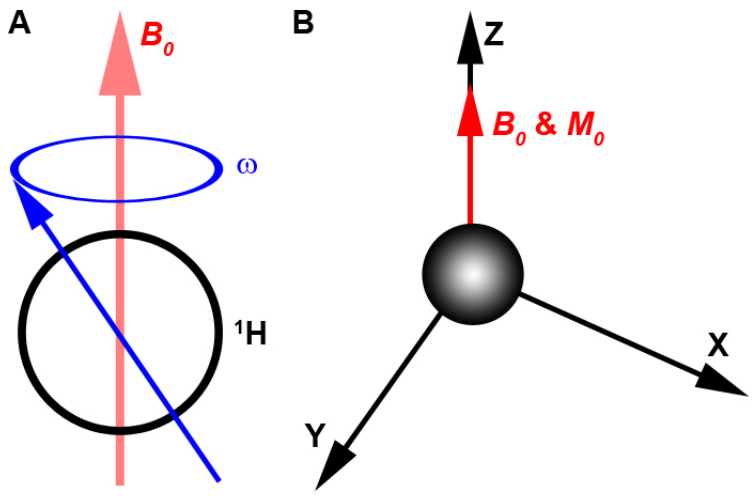

Figure 1.1: **Visualisation of bulk magnetisation**. **A**) Precession of bulk magnetisation around B_0 at frequency ω. **B**) Rotating frame of reference.

of the Bloch equation describing the motion of bulk magnetisation in a field. Before application of the pulse, quantum systems prefer to be in a low-energy state but there are only a small excess of low-energy protons since thermal energy induces transitions between the energy states. The relative population sizes of the lower and higher energy states can be determined using ΔE, Boltzmann's constant k and the temperature of the sample T. Increasing B_0 in turn increases ΔE, so the scanner can become more sensitive to population changes; hence scanners usually operate at high field strengths (greater than 0.5 Tesla) as T cannot be decreased for clinical purposes.

The initial bulk magnetisation M_0 has no vector component perpendicular to B_0, but it does have a component along B_0 (Fig 1.1B) due to the excess of the lower energy state nuclei. When M_0 is disturbed from this equilibrium state by a second magnetic field B_1 which is perpendicular to B_0 – called a radio frequency (RF) pulse – it will experience a turning force and precess at ω_L around B_0. The flip angle $\alpha = \gamma B_1 t_p$ describes the shift from M_0 by applying B_1 for t_p seconds. In structural imaging α is usually 90^o (flipping M onto the transverse axis) or 180^o (an inversion pulse). Once M has a component on the transverse axis, it generates its own rotating magnetic field which causes free induction signal to flow in a receiving coil. Free induction decay causes this signal to decay to zero as M relaxes back to M_0. Two types of relaxation simultaneously occur:

- Spin-lattice (T_1) relaxation which causes the precessing magnetic moments to align back with B_0 in the longitudinal direction M_z;

- Spin-spin (T_2) relaxation where the transverse component of M which is perpendicular to B_0, M_{xy}, is reduced to zero. Spins can also dephase due to field inhomogeneity, leading to T_2* relaxation.

Spatial localisation of the signal is achieved by using magnetic field gradients G (pictured in Fig 1.3) whose amplitudes vary linearly with position and add to the main field. At the magnet isocentre the field will be exactly B_0; elsewhere the frequency encoding, phase encoding and selective excitation (or slice selection) gradients will make the resonant frequency spatially variant. The thickness of 2D planar slices produced, stacked in the z direction, is $\Delta z = \frac{\Delta\omega}{\gamma G_z}$. $\Delta\omega$ is the bandwidth of frequencies in the slice selective pulse, relative to the full width at half maximum Δt of its frequency spectrum.

The MRI signal acquired from the output of a quadrature detector is sampled along the frequency encoding axis at each phase encoding step to produce a k-space representation in complex form (Fig 1.2A). An inverse discrete Fourier transform of the k-space data can then produce magnitude images (Fig 1.2B) such as structural T_2- and PD-weighted volumes. The phase information has several applications including flow velocity imaging in phase contrast angiography, phase-sensitive inversion recovery MRI [Ma, 2005] to enhance T_1 contrast, and for correcting geometry and intensity distortions in EPI through field mapping.

The resulting volume elements (voxels) can be considered the integral of signal over the voxel dimensions. A typical structural slice matrix size is 256×256 voxels; since the physical dimensions of the slice, or field of view (FOV), is usually 240mm^2 then to make roughly cubic voxels the slice thickness is often approximately 1mm. The voxel size may be increased in order to improve signal-to-noise ratio (SNR) with the through-slice resolution often sacrificed first.

1.1.2 Structural contrasts and pulse sequences

The intensity value at each voxel depends on the characteristics of the tissue at that location as well as the pulse sequence design (pictured in Fig 1.4), which can create T_1-, T_2- and proton density (PD)-weighted structural images. The defining features of each weighting factor are three parameters, shown in a pictorial pulse sequence representation in Fig 1.3:

- The repetition time (TR) denoting the time elapsed between two successive excitations of the same slice, which lets M return to M_0. Increasing the TR increases the T_2 contrast, imaging time and SNR, but reduces the T_1 weighting of tissue;

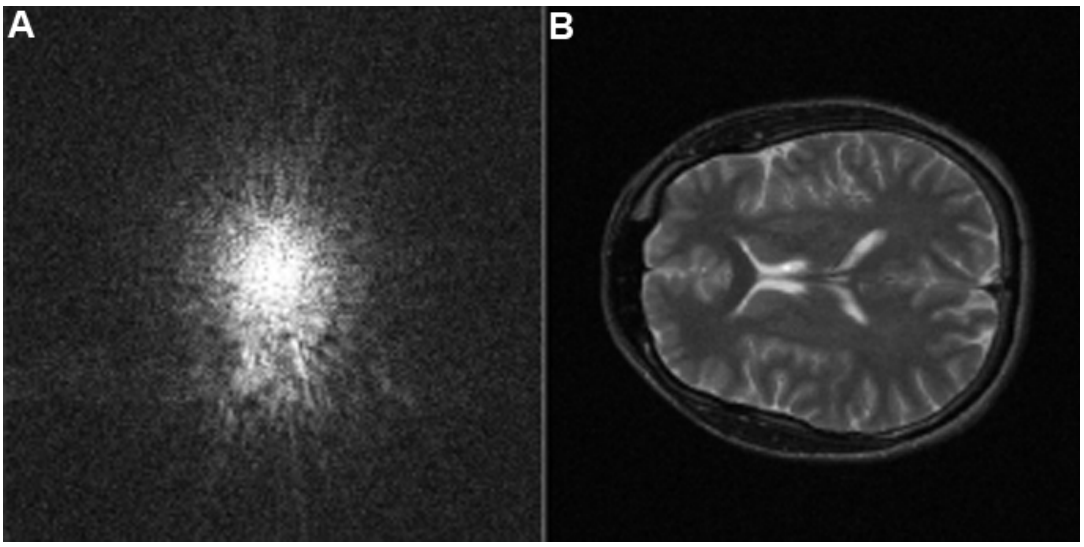

Figure 1.2: **Data representations in MRI**. Figure reproduced from `http://www.scopeonline.co.uk/pages/tutorials/kspace.shtml`. **A**) Magnitude of the complex-valued k-space representation of the signal. **B**) Magnitude image representation of the signal, following an inverse Fourier transform of the k-space representation.

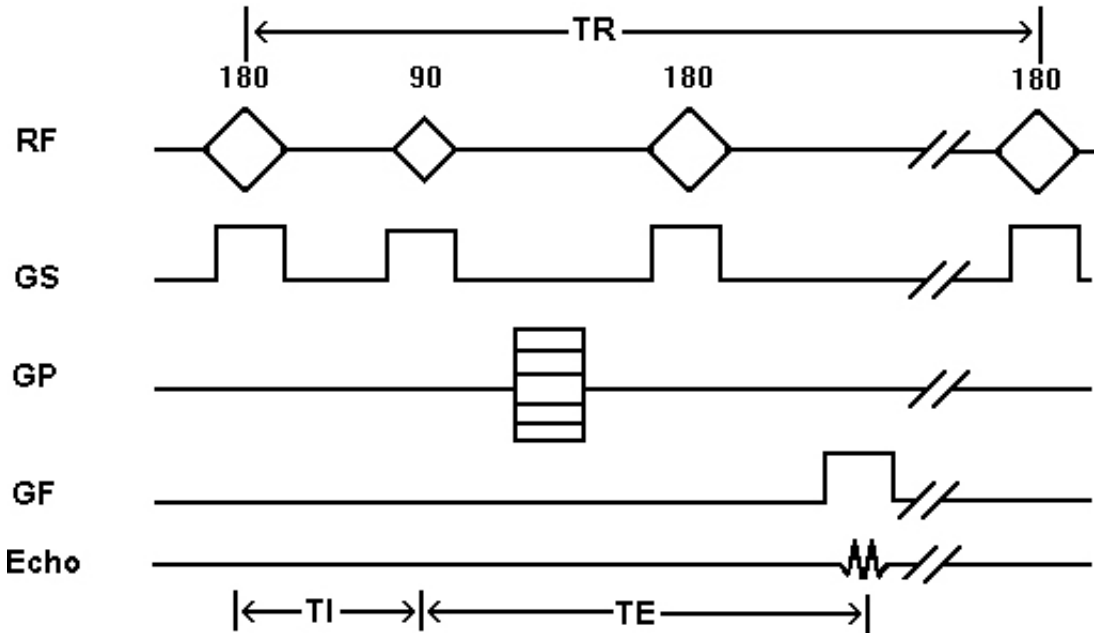

Figure 1.3: **Pulse sequence parameters**. Figure reproduced from `http://www.mritutor.org/mritutor/invers.htm`. A pictorial representation of a T_1 pulse sequence is shown and labelled with the repetition time (**TR**), echo time (**TE**) and inversion time (**TI**). **RF** = radio frequency pulses with labelled flip angles, **GS** = slice selective gradient, **GP** = phase encoding gradient, **GF** = frequency encoding gradient, **Echo** = signal received.

- The echo time (TE) measures the time between the RF pulse and the acquisition of the central frequency encoding sample. Decreasing the TE also decreases the T_2 weighting of tissue, but increases the T_1 weighting. Long TEs will result in reduced SNR as there is more time for spin dephasing and relaxation back to M_0 to occur;

- The inversion time (TI) is the length of the longitudinal magnetisation relaxation period between a 180^o inversion pulse to invert M and a following 90^o pulse, after which the signal can be measured on the transverse axis. The TI is not relevant to T_2 or PD sequences but can be chosen to suppress signal from certain tissues in inversion recovery (IR) sequences, such as for producing a tag image in arterial spin labelling (ASL) perfusion sequences.

In T_2- and PD-weighted imaging, spin echo (SE) sequences use a 180^o RF pulse to correct for M dephasing after an initial 90^o pulse. Gradient echo sequences (not used during testing in this thesis) are instead generated by using a pair of bipolar gradient pulses with a defined flip angle, but without the 180^o rephasing pulse then the decay of the transverse magnetization is more sensitivite to magnetic susceptibility artifacts. PD and T_2 contrasts can be acquired simultaneously using a dual-echo sequence by specifying two TEs: the first short TE (TE_1) is used to construct the PD volume, whereas the following longer TE (TE_2) creates the T_2 volume. Both volumes are perfectly co-registered – aligned to the same space – since they are acquired simultaneously.

In the neuroimaging community there have been reports of increased acquisition times, decreased SNR, and more image non-uniformity in dual-echo T_2-PD acquisitions compared to T_1-weighted volumes [Arnold et al., 2001]. Furthermore, contrast between gray and white matter tissues in T_1 may be more noticeable. However, studies acquiring brain volumes using several image weighting factors as well as multiple modalities are now commonplace (particularly in stroke imaging), not all tissues are contrasted well using T_1 (as investigated in Section 1.2.2), and improvements in scanner hardware have recently brought dual-echo sequences back into the research spotlight [Fiebach et al., 2004].

1.1.3 Noise, distortion and artifacts

In a homogeneous and static sample, noise in MRI is additive and Rician distributed. The Rician distribution possesses both spread and skewness parameters but can be

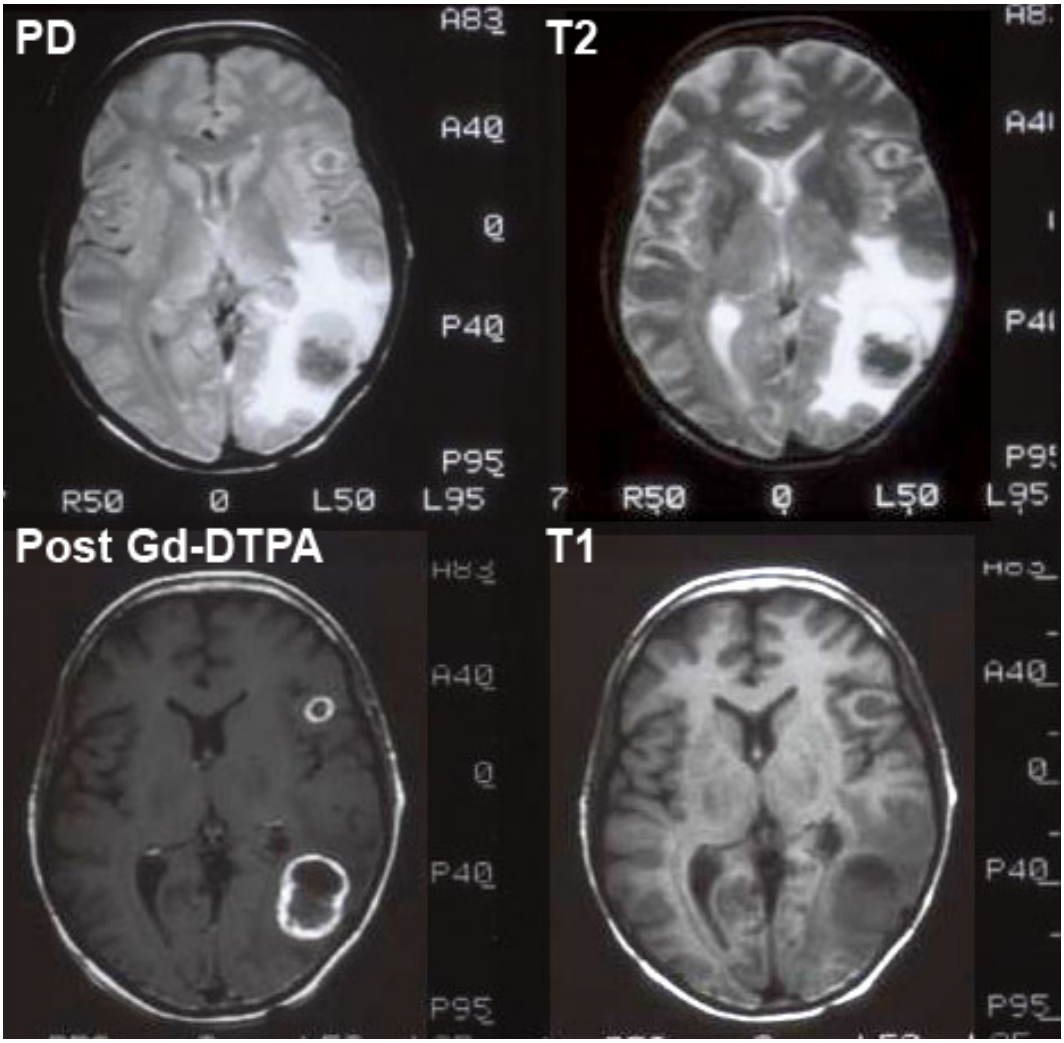

Figure 1.4: **Structural MRI weighting factors**. Figure reproduced from `http://www.rad.msu.edu/education/courseinfo/chm_domain/NMS_/JH/Default.htm`. Two lateral posterior and temporal metastatic tumours with surrounding vasogenic oedema are visible. **Post Gd-DTPA**: T_1 post-Gadolinium.

shaped similarly to a Gaussian. Many factors contribute to the intensity distributions for organic tissues including including material heterogeneity, field inhomogeneity and sample motion, and so Lyapunov's central limit theorem suggests a Gaussian distribution would be a suitable approximation. Furthermore, Gudbjartsson and Patz [Gudbjartsson and Patz, 1995] have observed that noise distributions are nearly Gaussian for SNR > 2. SNR is an important quality consideration, though above 20:1 the benefits are hard to observe and so reducing scan time or increasing the resolution would be more worthwhile. SNR is affected by:

- The number of signal averages (NEX) taken, where SNR is proportional to $\sqrt{\text{NEX}}$ and so the benefit of large numbers of signal averages is reduced. Scan time then increases linearly with TR $\times$ N_{PE} $\times$ N_{slices} $\times$ NEX, where N_{PE} is the number of phase encoding steps and N_{slices} is the number of slices;
- The voxel dimensions (through setting the slice thickness, matrix size and FOV), where the relative change in SNR is proportional to the fractional change in voxel size multiplied by the square root of the fractional change in N_{PE};
- The pulse sequence parameters and hardware-dependent factors including bandwidth, field strength and RF coil (B_1) sensitivity.

Multiplicative intensity distortions, collectively known as the bias field (Fig 1.5), vary slowly across the sample and can lead to reduction in the high frequency contents of the image (such as edges and contours) as well as spatial variance of tissue intensity statistics of more than 30% in a 1.5 Tesla scanner [Aubert-Broche et al., 1997]. While human vision can easily adapt internal models for tissue representation affected by low-level variation, many automated algorithms rely on assumptions of global intra-class intensity homogeneity. Various factors contribute to the formation of the bias field including lack of uniform sensitivity of the RF coils, gradient-induced eddy currents, magnetic susceptibility of tissue, attenuation of the RF signal inside the sample and inter-slice cross-talk [Simmons et al., 1994]. Since there is dependency on the sample properties as well as the pulse sequence parameters and current status of the scanning equipment, these effects cannot be eliminated by calibration alone and retrospective correction is required (either at the pre-processing stage, or during segmentation).

Artifacts are specific forms of distortion; many are hardware-related issues affecting MRI which are not encountered in this thesis. Movement of the sample leads to

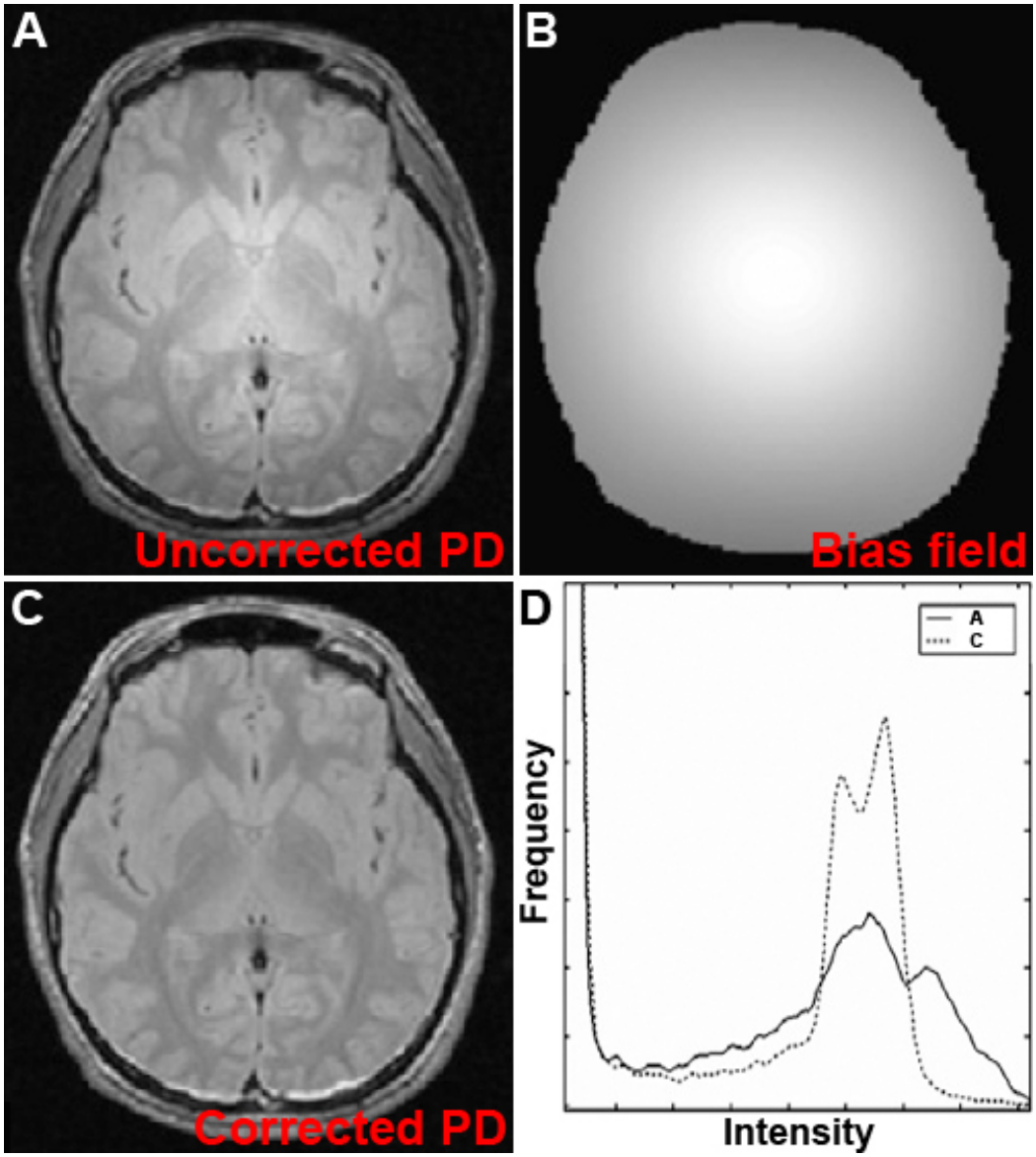

Figure 1.5: **Bias field effects and correction**. Figure taken from [Manjon et al., 2007]. **A**) Axial slice of a real PD volume corrupted by bias in the medial region. **B**) Estimated bias field of **A** using the method presented in [Manjon et al., 2007]. **C**) Corrected version of **A** using **B**. **D**) Changes to intensity histogram.

motion artifacts such as ringing, blurring or ghosting of the image which is very difficult to correct, and so in clinical brain imaging the patient's head is immobilised. It is also possible that slice selective gradients may not form perfectly rectangular blocks, and hence some interleaved slice acquisitions can lead to tissue experiencing repeated excitation. This issue can be alleviated using inter-slice gaps (at the penalty of not sampling some of the volume) or altering the sequence to ensure that the bulk magnetisation M has sufficiently relaxed for each slice.

Partial volume effect (PVE) occurs whenever tissues mix within the confines of a voxel – usually at structure boundaries, but also in striated tissue. Improving the resolution of the acquisition leads to a reduction in the relative volume of affected voxels (Fig 1.6). The intensity formation model most widely appreciated is a linear combination of the constituent pure tissues' intensities and their fractional contents (more fully covered in Section 5.1). PVE thus causes ambiguity in labelling decisions (Fig 1.15) concerning intensities on structure borders, as well as the softening of very thin structure intensities. Bias field effects and noise further complicate matters, and from a probabilistic modelling perspective, PVE causes the intensity histogram to become less sharp and increase in entropy.

1.1.4 Clinical structural brain imaging

A typical clinical MRI examination of a human brain will begin with staff checking the medical history, vital statistics and MRI safety record (contraindications) of the patient. While some subjects are able to tolerate an hour lying still inside the scanner, if the patient is restless, distressed, or suffers from claustrophobia, tremors, seizure or other involuntary movement disorder then the scan may have to be shortened or cancelled. Therefore time can be at a premium when considering SNR and resolution during the design of a pulse sequence, especially if several volumes are being acquired. Brain volumes acquired with several imaging modalities or with different MRI weighting factors prove extremely useful for pathology identification, providing anatomical context for low resolution scans, and capturing a range of contrasts. For example, the Lothian Birth Cohort 1936 (LBC1936) standard examination[2] includes diffusion tensor imaging (DTI), structural T_1, and dual-channel structural T_2-PD scans.

Once the patient has been prepared for scanning, a calibration scan takes place to mitigate bias and other confounds. The "birdcage" head coil is most commonly

[2]Performed at the Western General Hospital, Edinburgh, UK.

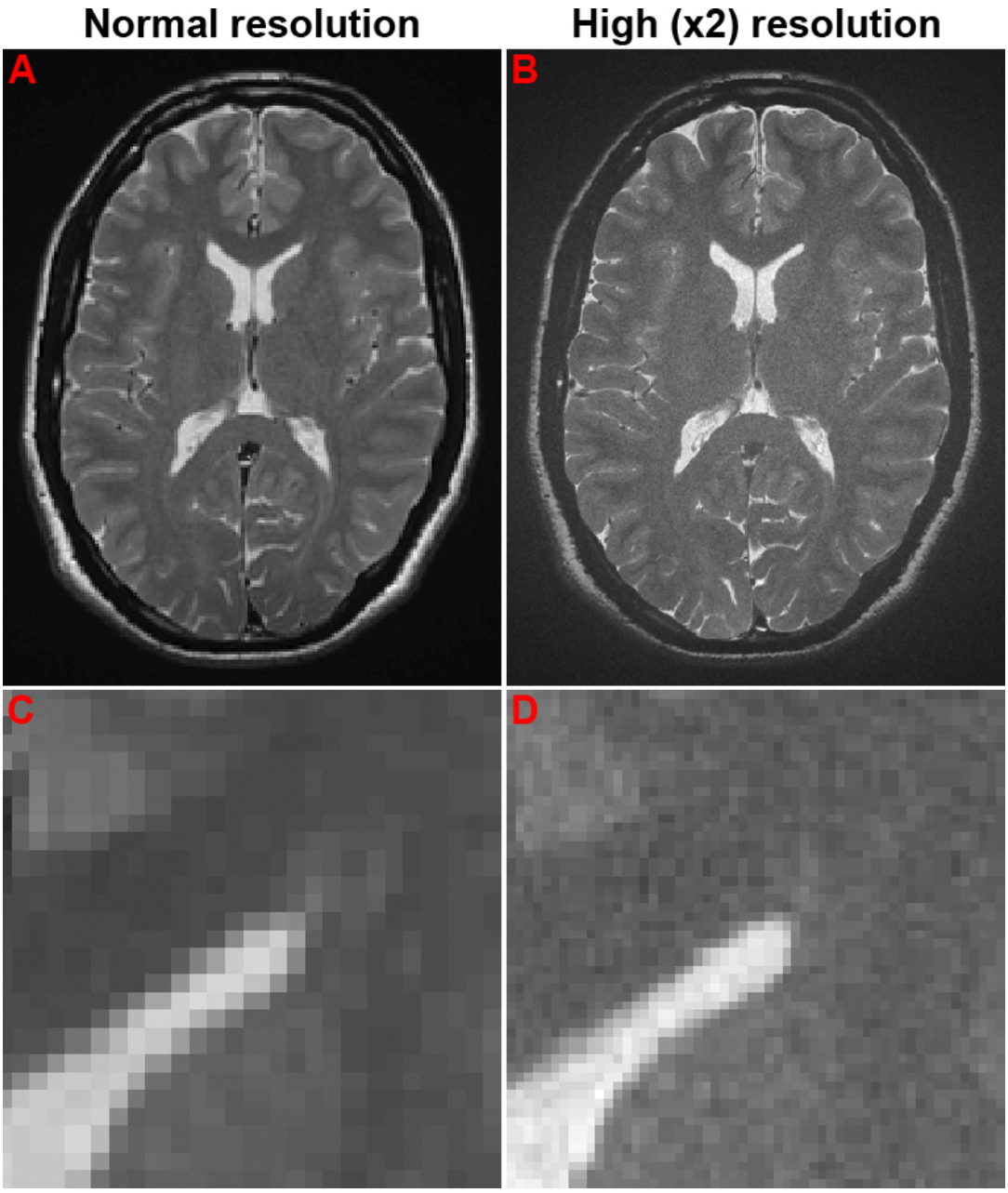

Figure 1.6: **Partial volume effect is increased by low resolution acquisition**. **Top row**: Whole T_2 slices. **Bottom row**: Close-ups of the top row images centred on the lateral ventricle. **Left**: Acquisition at 256×256 in-plane resolution with 2mm slice thickness. **Right**: Acquisition at 512×512 in-plane resolution with 1mm slice thickness. Slight differences between results stem from structures being covered slightly differently in the through-slice direction. Data was acquired at the Western General Hospital as detailed in Section 2.1.

used for scanning the brain. This coil creates the B_1 field and detects the FIS, and its small size allows safety margins to be more readily attained and a more uniform field generated. It also contributes towards a small point spread function (PSF) – generally less than a voxel's dimensions – for the imaging system so the voxel intensities are not considered to be composed of signal which originates from a wider locale. Prior to each sequence the region of interest (ROI) must be defined by a radiologist using a fast localising scan. The ROI is generally a cuboid shape, consisting of several slices which are stacked around a structure of interest. The 3D orientation and position of the ROI may need to be redefined for each sequence due to patient movement.

1.1.5 Temporal clinical brain imaging acquisition methods

Only structural imaging is considered throughout the bulk of this thesis, though two types of rapid imaging that can measure temporal changes are examined in the Conclusions (Chapter 6). Diffusion imaging is used to study diffusion: the random translational motion of the molecules caused by thermal energy. This term is also referenced throughout this thesis to describe a volume resolution enhancement process (Chapter 4), and the concepts of isotropy and anisotropy (Fig 1.8) are used to describe spatial filters which have preferences in orientation and extent (Section 3.4.4). The integrity of white matter tracts which normally display strong anisotropy of water flow can be investigated using diffusion imaging as well as brain abscesses and salvageable tissue post-stroke [Schlaug et al., 1997], and it can be used to grade the cellular swelling and cellularity of tumours.

Diffusion imaging studies the apparent diffusion co-efficient (ADC) since diffusion can also be caused by factors such as patient motion, bulk fluid flow and cellular transport. Ageing causes a mild but significant increase in the ADC, as do neurodegenerative diseases, due to white matter degeneration [Hergan et al., 2002]. Diffusion-weighted imaging (DWI) sequences can be sensitive to elevated T_2* and reduced ADC, whereas diffusion tensor imaging (DTI) sequences produce a positive symmetric tensor D which describes the local average molecular mobility along each spatial axis and the correlation between these directions. The principal dimensions of a diffusion ellipsoid correspond to the eigensystem of D, and its anisotropy can usually be described as linear (one large eigenvalue, Fig 1.8A) or planar (two large eigenvalues, Fig 1.8B). DTI produces three measures at each voxel:

- The mean diffusivity characterises the displacement of molecules;

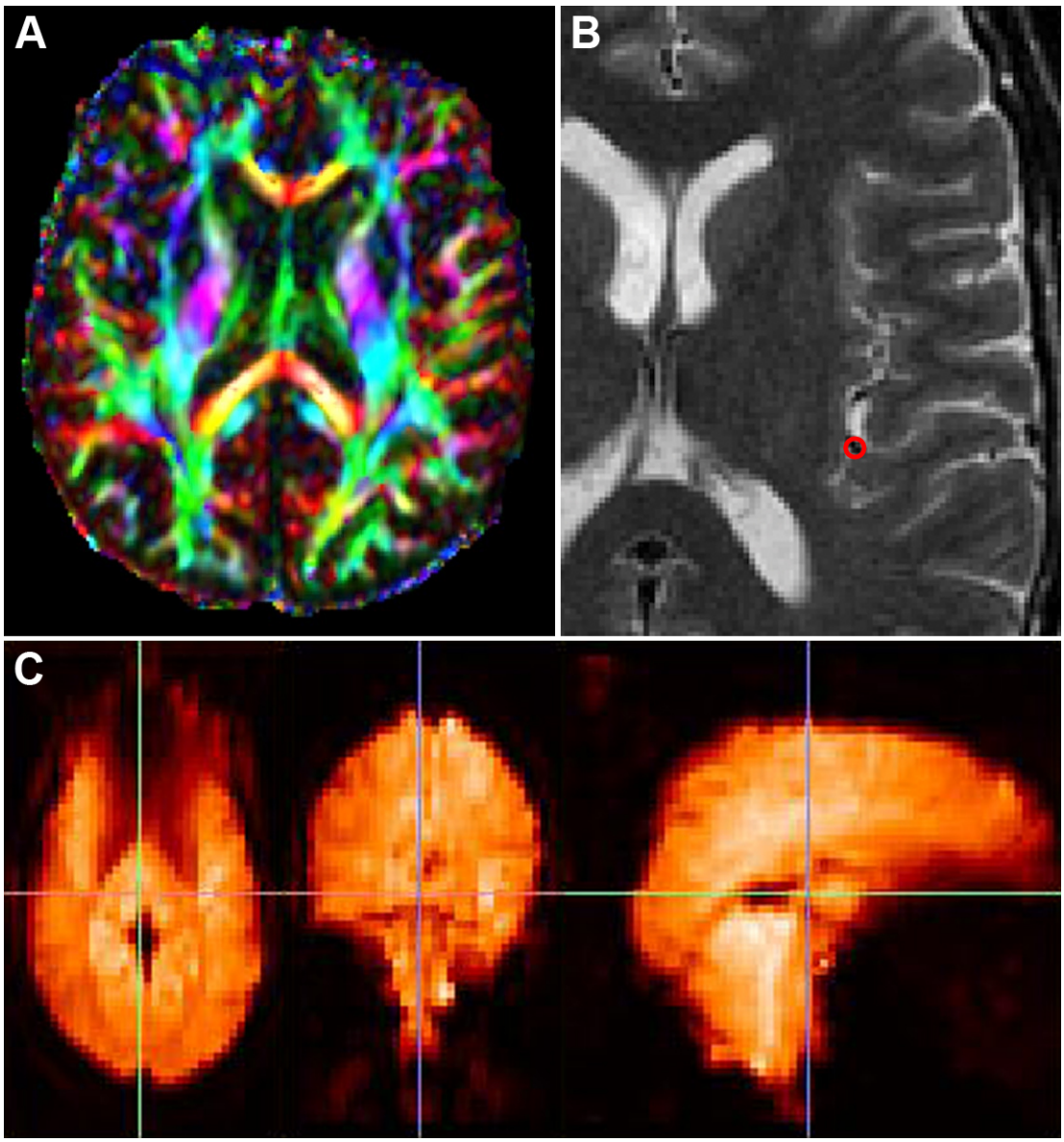

Figure 1.7: **Temporal brain imaging**. **A**) Axial slice of a fractional anisotropy (FA) map produced by diffusion tensor imaging (DTI). Colours correspond to the principal direction and relative strength of maximal diffusion. Figure reproduced from `http://www.loni.ucla.edu/~thompson/MATH/math.html`. **B**) Region of interest (ROI) marked in red to define an arterial input function (AIF) for a temporal region of the brain. T_2 data acquired from the LBC1936 dataset. **C**) Raw functional MRI (fMRI) data acquired at 128×128 in-plane resolution. Axial, coronal and sagittal planes are shown from left to right. Figure reproduced from `http://brainimaging.waisman.wisc.edu/~oakes/spm/visual_stim_demo/fmri_visual_stim.html`.

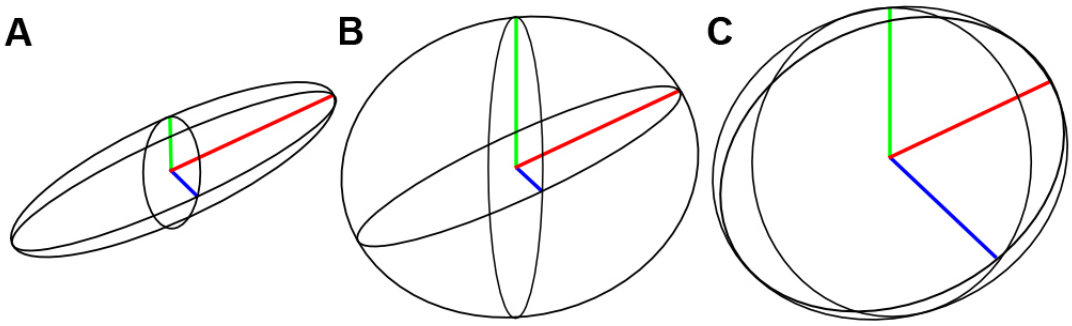

Figure 1.8: **Diffusion anisotropy illustrated by prototype ellipses**. Coloured lines indicate the principal axes of the ellipses. **A**) For an anisotropic linear diffusion ellipsoid, the statistical average of flow is biased towards a certain direction. **B**) It is biased along a certain plane for anisotropic planar diffusion. **C**) There is no bias for an isotropic diffusion spheroid so each principal axis is a similar length.

- The fractional anisotropy (FA) measures the fraction of the magnitude of D that can be attributed to anisotropic diffusion (Fig 1.7A);
- The principal direction of anisotropy which is formed by the eigenvector of D with the largest eigenvalue (Fig 1.7A). Some DTI methods (such as Q-ball imaging) may be able to extract several directions i.e. in regions with crossing white matter fibres.

Perfusion imaging can measure the blood supply of tissue using an intravenous contrast agent such as gadolinium using a dynamic susceptibility contrast sequence, or non-invasively with reduced SNR using arterial spin labelling. Cerebral blood flow, cerebral blood volume, mean transit time and time-to-peak measurements can be produced which are relative to an arterial input function (AIF) ROI placed over a local blood supply (Fig 1.7B). These measures can distinguish ischaemic and necrotic tissue from healthy tissue, detect haemorrhaging, and the mismatch between CBF and CBV can be used to grade tumour severity.

1.1.6 MRI phantoms

Phantoms are customisable constructs used to calibrate the performance of MRI-related systems. The material content and positions of material regions inside are constant and precisely known and hence the true labelling, intensities or other parameters can be easily calculated. Physical phantoms[3] are objects built using MRI-compatible materials

[3] Also known as homogeneity phantoms or resolution phantoms.

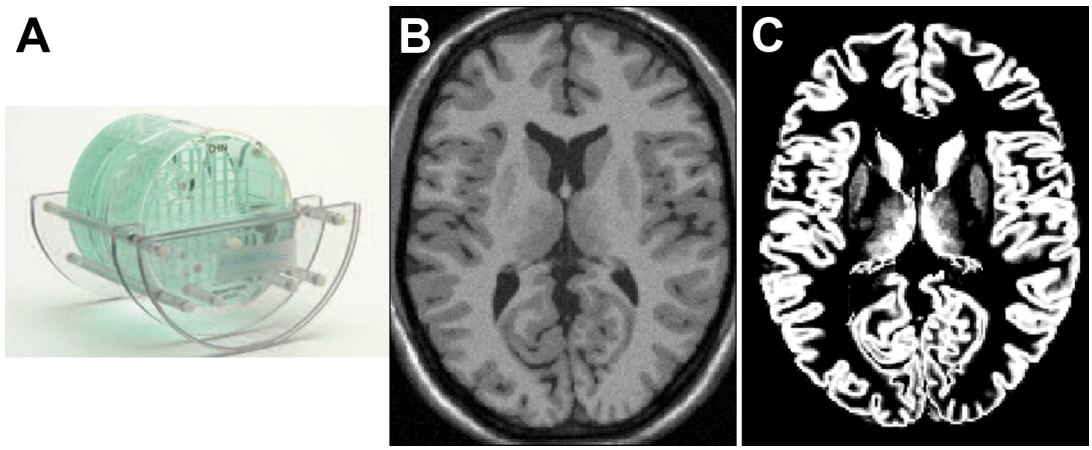

Figure 1.9: **Phantoms**. **A**) Physical water-based plastic-cased phantom. Picture reproduced from `http://www.newmaticsound.com/index.php?app=ccp0&ns=prodshow&ref=ACRPC`. **B**) Axial T_1 slice from the BrainWeb [Cocosco et al., 1997] software phantom with 3% noise added. **C**) Gray matter tissue content map which generates **B**.

with known T_1, T_2, and PD values (Fig 1.9A). Rearrangeable sections can create different testing configurations, and they are immobile to offer ideal imaging conditions (which will not be realised in clinical practice). They are usually scanned to measure the contrast-to-noise ratio (CNR) and spatial uniformity of the RF coil B_1, as well as to test spatial properties of the scanner including the resolution and SNR as a function of position. Physical phantoms have not been used in this project, but they are addressed in suggestions for further testing (Chapter 6).

Software phantoms provide a simulation of MRI acquisition as well as a model of the sample properties and structure to test the robustness of MRI data processing methods to different levels of artifacts, contrast and noise. BrainWeb [Cocosco et al., 1997; Collins et al., 1998] is an advanced brain phantom (Fig 1.9B) which has been used to test a large number of segmentation algorithms and also features in this thesis. The phantom's construction was based on a semi-automated classification of 27 registered acquisitions of the same subject. Each voxel's intensity value was averaged over the acquisitions, and the T_1-weighted gradient echo acquisition possessed 1mm isotropic voxels, low noise, TR = 18ms, TE = 10ms and a flip angle of 30 degrees. Importantly, since BrainWeb volumes with different MRI weighting factors are generated using the same tissue model (Fig 1.9C) then they are all perfectly co-registered. The features of BrainWeb include:

- Production of T_1-, T_2- or PD-weighted intensities;
- Different voxel sizes and modelling of tissue mixing inside voxels;
- A range of noise levels can be added;
- A smoothly-varying bias field of varying severity can be induced;
- A somewhat realistic model of brain anatomy, allowing testing on data similar in size and appearance to real human brain volumes. However, no blood vessels are included in the model and so analyses concerning this class cannot be performed. More extensive commentary on the realism of the simulation is presented in the Conclusions (Chapter 6);
- A volume with pathological tissue affected by multiple sclerosis can also be selected.

1.2 Brain anatomy

In this section a rudimentary outline[4] of brain anatomy is presented to provide structural context for MRI volume data and segmentation results. The description takes place at several levels, from gross feature types down to individual structures and details of the constituent tissue types. Illustrations of many of the structures have been provided[5] so that their structural and topological peculiarities can be appreciated.

1.2.1 Directions and referencing

Three planar orientations (Fig 1.10B) are commonly used to view 2D images of 3D brain volumes: axial (on which data across the X and Y axes are shown), sagittal (Y-Z) and coronal (X-Z). The X axis is also called the left-right or medial-lateral axis; the Y axis can be named anterior-posterior, rostral-caudal, or front-back; and up-down, superior-inferior or dorsal-ventral are terms for the Z axis (Fig 1.10A).

Regions of the brain can be referred to using numbered Brodmann's areas, though it has been argued that the sulcal landmarks used may not accurately demarcate cytoarchitectural boundaries [Zilles et al., 1997]. Stereotaxic Talairach atlas or Montreal

[4]For a more detailed description of brain anatomy, and details of the variations in shape and configuration between individuals, the reader is referred to an excellent work by Crossman and Neary [Crossman and Neary, 2000].

[5]Taken from Gray's Anatomy [Gray, 1918].

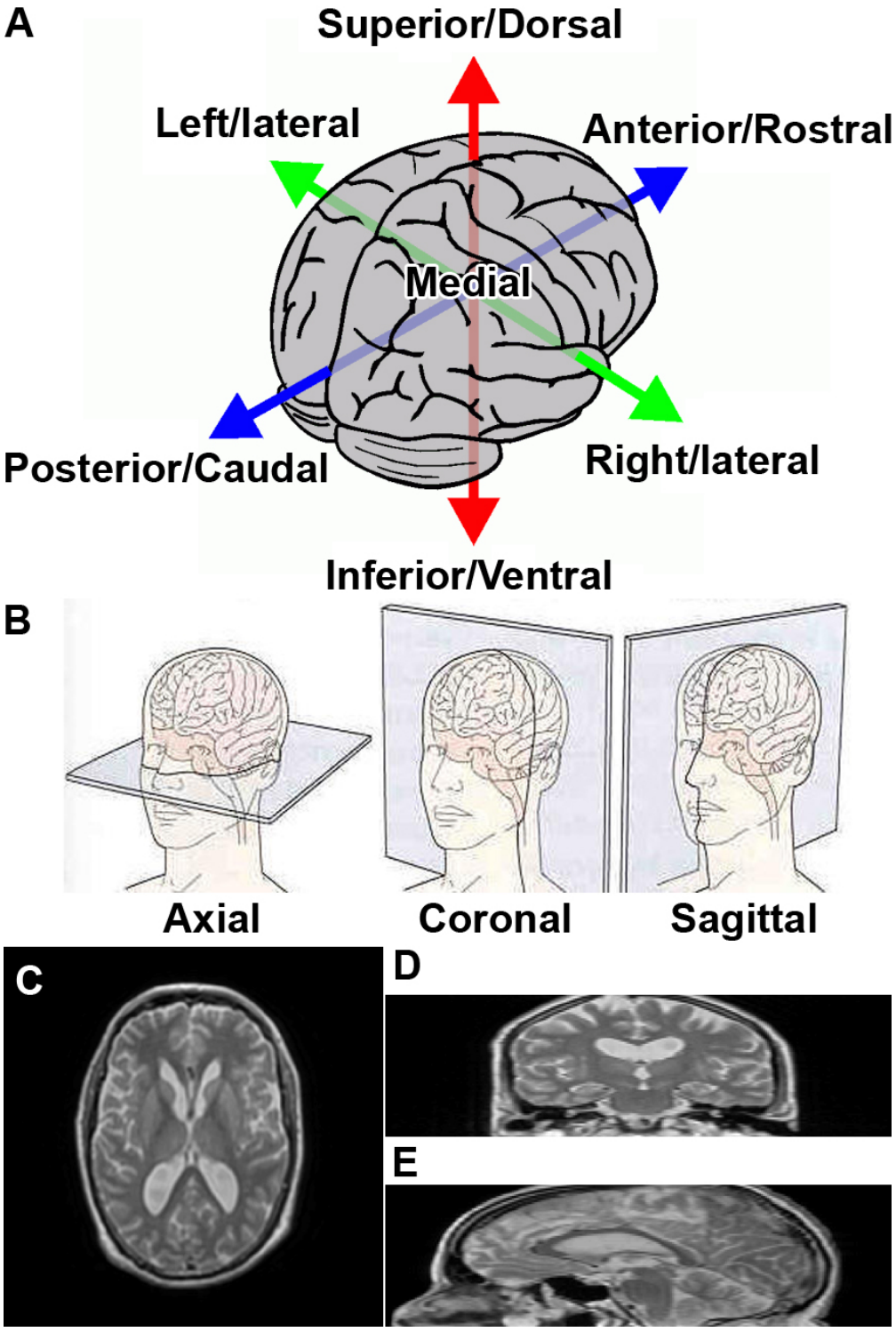

Figure 1.10: **Directions and referencing**. **A**) Anatomical referencing directions. **B**) Slice axis names; left to right these are axial, coronal and sagittal. Illustration reproduced from `http://serendip.brynmawr.edu/bb/kinser/BrainInfo.html`. **C**) Axial slice from a T_2 volume in the LBC1936 dataset. **D**) Coronal slice taken from the same volume as **C**. **E**) Sagittal slice taken from the same volume as **C**.

Neurological Institute standard brain co-ordinates can also be used but these atlases are produced by averages of brain volumes which are transformed to the same space, or are prototypical brains representing some population. Since brains contain many different configurations and numbers of features between individuals – for example the smaller cortical sulci (Section 1.2.3) – then the registration of the target volume to the atlas may be problematic. Therefore, in this thesis the names of the anatomical structures have been used to reference them.

1.2.2 Brain tissue types

The functional content of the brain (parenchyma) contains five main tissue types, and Fig 1.11 shows the relative intensities of each in the main structural imaging weighting factors. The edges of the brain parenchyma are adherent to the meningeal layers; further outside are the skull, skin, hair, muscles and organs of the head, and finally the surrounding air which has an intensity close to zero in MRI (Fig 1.12B) – a large proportion of MRI images are made up of such non-brain regions.

- The most numerous and widespread are the neuroglial cells[6] supporting the other tissues, which are not directly observable using MRI;
- White matter is the next most prevalent and consists of nerve fibres (axons) which connect neurons;
- Gray matter is enriched in neuronal cell bodies, and forms the convoluted sheet-like cortex at the surface of the brain as well as more blobby sub-cortical nuclei (also known as deep gray matter);
- Cerebrospinal fluid (CSF) is contained within the meninges as well as filling the ventricular system in the centre of the brain and approaching the spine. Water is a major component of CSF so it has a very high proton density;
- Blood flowing within arterial and venal vessels. The paramagnetic property of haemoglobin combined with rapid flow causes blood to have very low intensities.

[6]Usually shortened to glia.

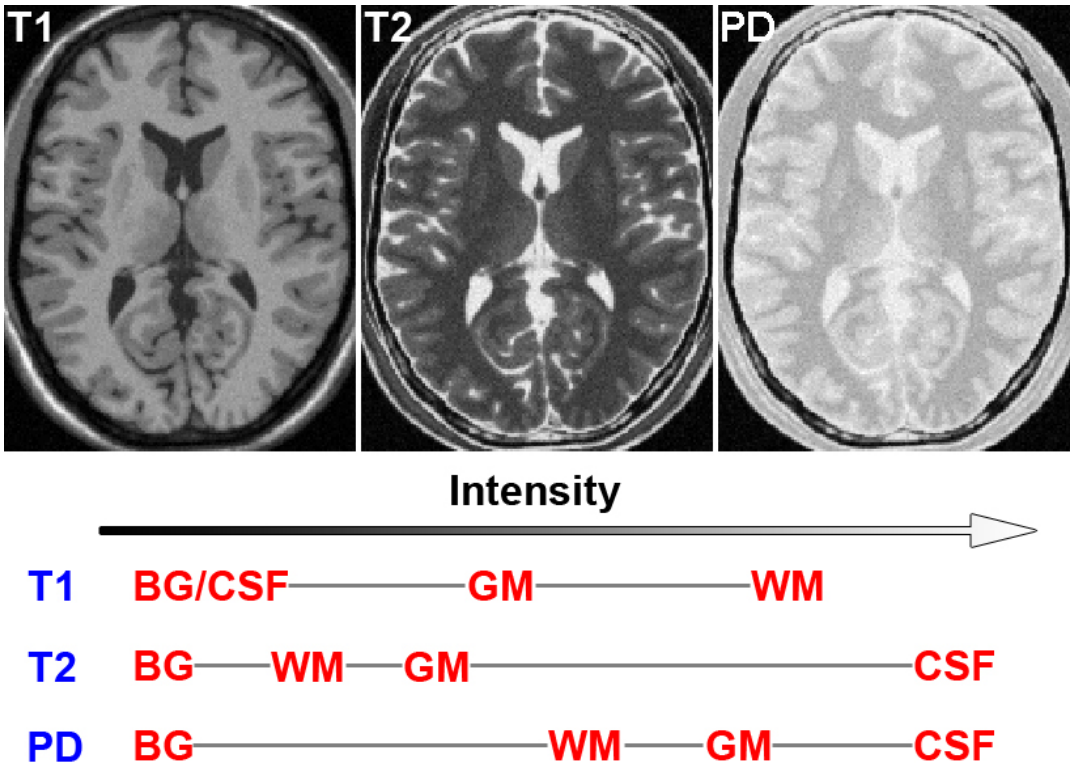

Figure 1.11: **Appearance of brain tissue types in the three main structural MRI weighting factors**. Data from the BrainWeb phantom with 3% added noise are shown. **A**) T_1 slice. White matter is the brightest, then gray matter, then blood, CSF and background. **B**) T_2 slice. CSF is brightest, followed by gray matter and then white matter, and finally blood and background. **C**) PD slice. CSF is brightest, followed closely by gray matter, then white matter, and finally blood and background. **D**) Relative values of tissue intensities.

1.2.3 Gross structure

The cerebrum comprises two hemispheres, which are incompletely separated by the great longitudinal fissure containing the falx cerebri which is a thin fold of dura mater. This is one of three membrane layers known as the meninges, which ensheath the brain within the skull (Fig 1.12B) but are not of great interest in segmentation since they are not part of the parenchyma. Each hemisphere contains an outer layer of cortex surrounded by the meningeal layers, an inner core of white matter, and sub-cortical structures and the ventricles. Inferior to the cerebrum lies the cerebellum, which also has two hemispheres; both the cerebellum and cerebrum are attached to the brainstem leading to the spinal cord. The cerebrum is divided into four lobes (Fig 1.12A), with the most extreme tips being known as poles:

- The frontal lobe is found in anterior part of the brain, containing the pre-frontal cortex at the very front and the pre-motor and motor areas (pre-central gyrus) toward the parietal lobe;
- The parietal lobe is located posterior to the frontal lobe. It is divided from the frontal lobe by the central sulcus, a single continuous furrow running over the entire lateral surface of the hemisphere from the great longitudinal fissure to the lateral fissure;
- The temporal lobe is located bilaterally, around the level of the ears. It is separated from the parietal lobe by the lateral fissure, adjacent to which is the primary auditory cortex;
- The occipital lobe is the most posterior part of the brain, containing the primary visual cortex (also known as striate cortex or V1) surrounding the calcarine sulcus. The boundaries with the parietal and temporal lobes are indistinct, except the parieto-occipital sulcus at the medial surface.

CSF is present in furrows called sulci[7] between convolutions of the cortex known as gyri[8]. On the lateral surface of each cerebral hemisphere lies the largest sulcus, the lateral fissure, which divides the cortex into four lobes; within its depths is cortex known as insula. Cortical folding may induce rapid changes in the direction of sulci, and particularly tight folding may create sulcal interruptions: apparent breaks in sulci

[7] Singular is sulcus.
[8] Singular is gyrus.

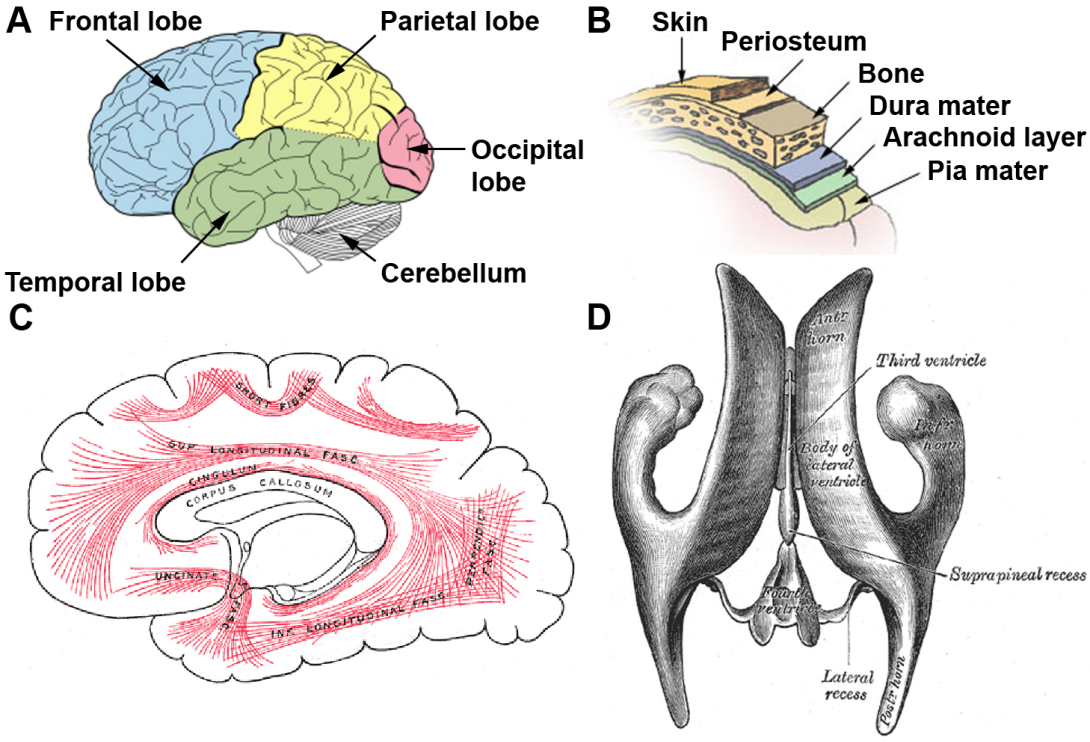

Figure 1.12: **Gross brain structure**. **A**) Cerebral lobes and cerebellum. **B**) Extra-parenchymal tissue, including the meningeal layers. Illustration reproduced from `http://en.wikipedia.org/wiki/File:Illu_meninges.jpg`. **C**) Ventricular system. **D**) White matter structures. Illustrations **C** and **D** were reproduced from Gray's Anatomy [Gray, 1918].

due to the convexity of the cortex. In [Ono et al., 1990] it was shown that some larger sulci are consistently located between human brains, and even between species, but the variations in both configuration and presence of smaller secondary and tertiary sulci are such that the pattern of cortical folding can be considered unique between individuals. CSF is produced by the choroid plexus, an invagination of the vascular pia mater into the ventricles, which are also joined to the sub-arachnoid meningeal layer. The ventricular system (Fig 1.12C) comprises:

- The fourth ventricle, arising from the central canal of the spinal cord, forming a rhomboid shape on the surface of the brainstem and beneath the cerebellum;
- The slit-like third ventricle, located between the thalamus and hypothalamus;
- The lateral ventricles, extensive curved chambers located in both hemispheres, with horns extending into the frontal, occipital and temporal lobes. The septum pellucidum separates the anterior horns, and adjacent structures include the caudate nucleus to its side, the corpus callosum forming a roof above it, and the thalamus and hippocampus.

White matter (Fig 1.12D) tracts are axons sharing the same connections and functions which run together and follow the same course. These are classified depending on their origin and destination:

- Projection fibres connect cortex and sub-cortical structures; one of the most prominent is the internal capsule, a broad sheet forming a fan-like arrangement (corona radiata);
- Commissural fibres run from one hemisphere to another. The most significant is the corpus callosum which is the largest uniform white matter structure in the brain at the base of the great longitudinal fissure and falx cerebri;
- Association fibres connect structures within the same hemisphere. The fasiculi connect the lobes and the cingulum connects the frontal and parietal lobes with hippocampal and temporal gyri.

The brain has no fat stores and so its blood supply (Fig 1.13E) is pervasive and the vessel network is extremely interconnected with redundant channels. The internal carotid and vertebral arteries form the main supply and lead to the Circle of Willis (Fig 1.13B) at the base of the brain. The Circle of Willis has a large number of possible

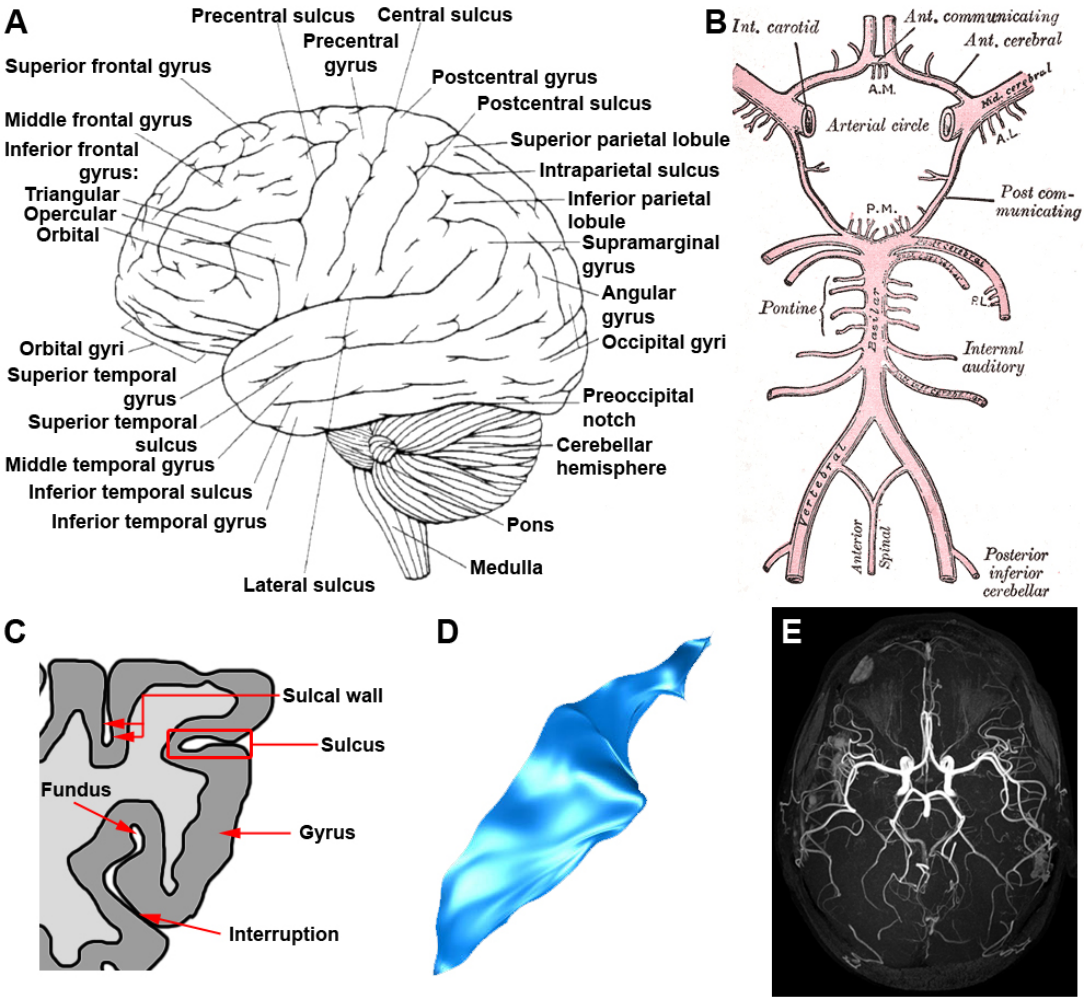

Figure 1.13: **Thin brain structures**. **A**) Larger cortical sulci and gyri. **B**) Circle of Willis. **AL** = Antero-lateral. **AM** = Antero-medial. **PL** = Postero-lateral. **PM** = Postero-medial ganglionic branches. Illustration reproduced from Gray's Anatomy [Gray, 1918]. **C**) Scale-scale structure of cortical sulci and gyri. **D**) 3D rendering of an extracted central sulcus. Figure reproduced from `http://sim3.univ-rennes1.fr/theme2/theme2gb.html`. **E**) Magnetic resonance angiography (MRA) projection revealing part of the blood vessel network. Image taken from `http://wsunews.wsu.edu/pages/publications.asp?Action=Detail&PublicationID=10646`.

configurations, and like sulci, the vessel network can also be considered unique between individuals. The venous drainage of blood is undertaken by superficial veins, deep veins and dural venous sinuses, of which the latter are the largest. The much smaller capillaries are not visible using MRI at clinically-viable resolutions. Compared to sulci, vessels turn less sharply and tend to be more consistent in thickness along their lengths except at junction points (also known as bifurcations).

Ageing typically leads to enlargement of the ventricles, widening of the cortical sulci and reduction in gray matter volume [Good et al., 2001] and widening of venous cavities. Lesions, blood vessel damage or stenosis, and cellular damage can result in oedema which has a similar appearance as CSF. In aged brains CSF-filled Virchow-Robin (VR) perivascular spaces become more common, surrounding small arterial vessels.

1.2.4 Gray matter structures

Cerebral cortex (or neocortex) forms the surface of each cerebral hemisphere; it is highly convoluted in order to maximise the surface area within the skull, and about 70% of the cortical surface is hidden within the depths of sulci. The development of gyri and sulci are linked and so the cortical folding pattern can also be considered unique between individuals. The cortex is asymmetric and the cortical thickness is variable, with anterior regions typically thicker than the most posterior ones [Fischl and Dale, 2000].

Within the cerebrum lie sub-cortical nuclei (Fig 1.14B). At the lateral walls of the third ventricle – in the dorsal to ventral direction – the diencephalon consists of the small epithalamus, large ellipsoid thalamus (containing internal medullary lamina of white matter), subthalamus and hypothalamus. Several large nuclei called the basal ganglia lie medial and lateral to the internal capsule, which itself lies between the caudate nucleus on the wall of the lateral ventricle and the putamen lateral to the internal capsule. Medial to the putamen is the globus pallidus. The thalamus and caudate are separated by white matter known as the stria terminalis.

The limbic system, on the medial rim of the brain, includes the amygdala near the temporal pole and next to the inferior horn of the lateral ventricle, the hippocampus, and several structures next to the inferior horn of the lateral ventricle: the septum, the cingulate gyrus, and the parahippocampal gyrus. White matter linking the limbic system includes the fornix and fimbria.

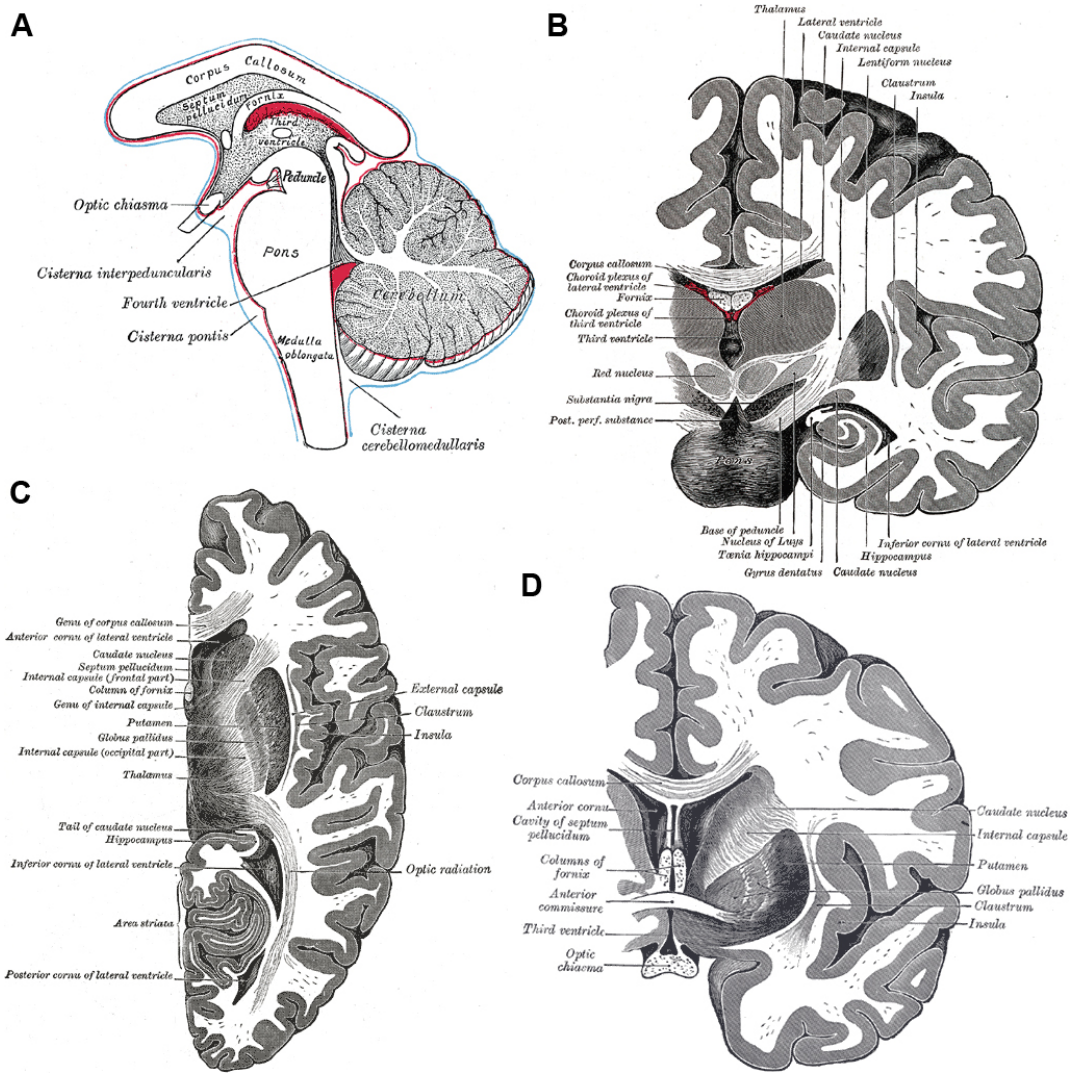

Figure 1.14: **Gray matter structures**. Illustrations taken from Gray's Anatomy [Gray, 1918]. **A**) Brainstem and cerebellum. **B**) Coronal section of the cerebrum immediately in front of the pons. **C**) Axial section of the right cerebral hemisphere. **D**) Coronal section of the cerebrum through the anterior commissure.

The cerebellum (Fig 1.14A) possesses an outer layer of cortex surrounding a central core of white matter and consists of two hemispheres joined in the midline by the vermis, similar to the cerebrum. The cortical surface is more complex causing patterns of narrow parallel folds or folia, giving the white matter a characteristic tree-like appearance. Within the white matter are four pairs of very small cerebellar nuclei. Cerebellar peduncles, lying lateral to the fourth ventricle, connect the cerebellum to the brainstem.

1.3 Brain volume segmentation

In a general image processing context, segmentation is the partitioning of data into non-overlapping coherent regions where the constituent elements have similar properties. Over-segmentation fractures the data into more regions than is correct, and under-segmentation accordingly produces less. The labelling of those regions produces a classification. Advanced automated segmentation algorithms attempt to aggregate, sub-divide or re-model those regions to produce a more accurate classification without expert assistance. Hard segmentation refers to labelling the data elements as each belonging to a single class; soft segmentation quantifies the normalised fractional content of each class.

In this section the instance of MRI brain tissue segmentation[9] using volumes acquired with several image weighting factors is introduced with details of the necessary pre-processing to help ensure a satisfactory result. The current issues in the field have also been discussed in the context of the deviation of a widely-used segmentation tool's results from a human expert's labelling in order to generate the goals of this thesis.

1.3.1 Segmentation basics

In clinical imaging, segmentation is used for brain tissue volume quantification such as analysis of brain asymmetry [Herbert et al., 2005], calculation of lesion or tumour volumes [Prastawa et al., 2003], and studies of white matter degeneration, gray matter thinning and ventricle enlargement [Good et al., 2001] in ageing. As a pre-processing step segmentation can assist manual labelling of the volumes and also provide anatomical context for other (generally lower-resolution) imaging types such as DTI, functional MRI, PET and SPECT. Assisted surgical planning and navigation depend upon

[9]For a more in-depth review of brain MRI segmentation methods, the reader is referred to [Pham et al., 2000].

the quality of segmented volumes [Grimson et al., 1996], as do other 3D visualisation applications (particularly for sub-cortical structure) [Bullitt and Aylward, 2002] and in structure analysis, quantifying modes of shape and texture variation across populations [Babalola et al., 2008].

The features used in MRI brain tissue segmentation are usually the volume intensities (multi-valued where volumes with several image weighting factors have been acquired) or their derivatives. Processing can be simplified through the consideration of only a few classes, and in the literature BG and more commonly *other* (comprising non-parenchymal tissue, or tissue that cannot be represented by the main brain parenchymal tissue types identified in Section 1.2.2 listed below) have not always been included due to their small membership in the brain parenchyma.

1. **BG**: a class combining background, the skull and blood, since these tissue types possess similar intensities in the main structural image weighting factors;
2. **WM**: white matter;
3. **GM**: gray matter;
4. **CSF**: cerebrospinal fluid.

It is normal to consider two-class mixing models since at clinical resolutions it is rare for a voxel to contain more than two tissue types. Voxels are either pure and contain one tissue type only, or are partial volumes and contain two. The mixel model is the most popular tissue mixing model (examined in more detail in Section 5.1), which proposes that voxel intensities are formed by a linear combination of representative class intensities and their respective fractional proportions (taken from the voxel labelling), with an additional noise component. Statistical methods for segmentation use probability distributions of voxel intensities (typically normal Gaussians, as investigated in Section 1.1.3) rather than simple thresholding to infer the most likely labels for the noisy data. However, PVE leads to ambiguity in segmentation based on intensity alone (Fig 1.15), even without taking into account the effects of noise and bias: it is possible that many of these linear combinations can produce the same intensity.

The piecewise-constant property suggested of MRI brain data proposes that the intensities and labels of structures can be considered locally constant, which is valid throughout a great proportion of the brain volume except where thin structures are insufficiently sampled, including at gradiations of intensity (around some sub-cortical

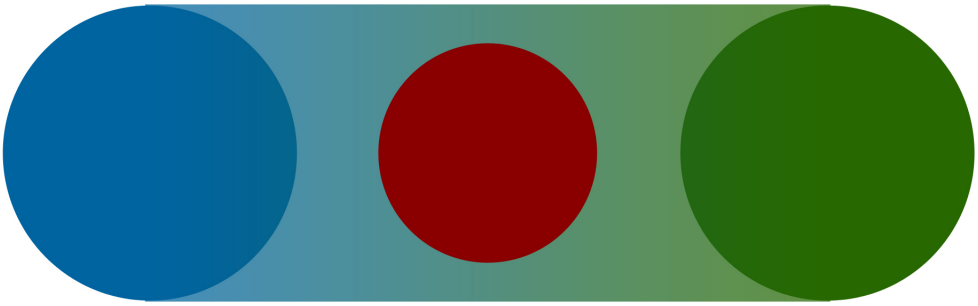

Figure 1.15: **Labelling of intensities can be made ambiguous by PVE**. A representation of a dual-channel intensity scatter plot (in a 2D intensity space) is shown. When there are at least three tissue classes, a pure class (red) may be confused with a mixture of two others (blue and green) due to their similar intensities.

structures such as the thalamus at typical clinical resolutions). Region growing approaches can be used to enforce this property and augment statistical segmentation methods but they require seed points and can be sensitive to bias, noise and PVE. Inventive approaches involving automated competing agents and spatially-distributed model estimation [Germond et al., 1999] to enforce the piecewise-constant property have also been proposed. More commonly, the context of a local neighbourhood can be considered while enforcing spatial coherence of labelling and reducing noise in large tissue regions. The connectedness of a neighbourhood system refers to how many other voxels are in the same neighbourhood as the central voxel. The Markov property states that contextual constraints can be determined solely by small neighbourhood cliques (Fig 1.16A), and this is implemented in 3D using Markov random field (MRF) models with the greedy iterated conditional modes (ICM) algorithm [Besag, 1974] usually employed as a time-efficient iterative optimiser to disambiguate which label may be most appropriate at each voxel.

Clustering is an alternative approach which assigns to data points membership in clusters, in order to minimise the feature distance between elements of the same cluster and maximise the distinctiveness of clusters. Each cluster possesses a centroid or prototype (in the k-means method) or it can be defined using sets of boundary cuts (spectral clustering). Fuzzy c-means (FCM) clustering is an extension of the k-means method which assigns a single data point a fuzzy normal membership value in all clusters. Clustering is a good option for initialising class parameters for more advanced algorithms since both the distance function (or adjacency matrix) between data points and the neighbourhood interaction models tend to be fast and simple, and the optimi-

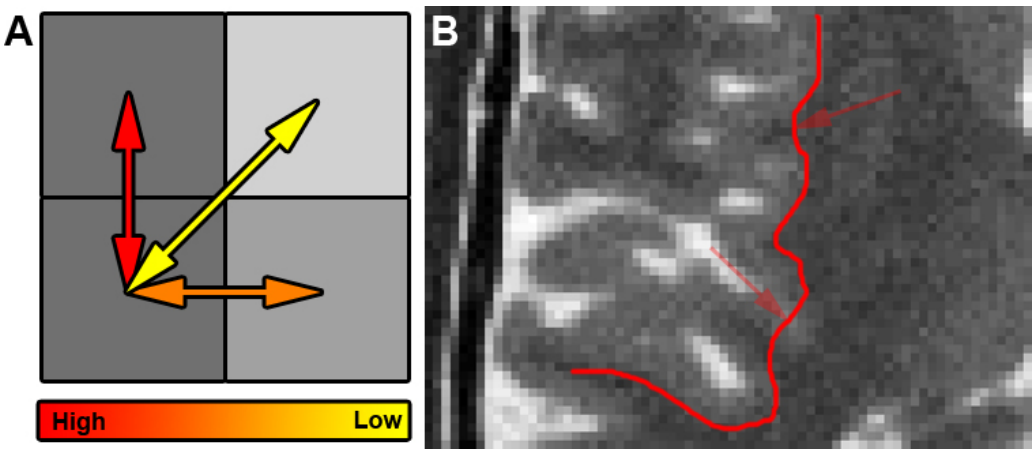

Figure 1.16: **Implementation of spatial dependence**. **A**) Scoring of MRF cliques; the pixel shading denotes the label, and the colour of the arrow denotes the strength of the interaction. **B**) Front propagation. Arrows denote the future progression of the red border to more closely match the white matter/gray matter boundary on a T_2 axial slice of the BrainWeb [Cocosco et al., 1997] phantom.

sation is often robust and easy to derive using Lagrange multipliers.

1.3.2 Registration

Co-registration (or fusion) of the volumes of the same subject (acquired with different modalities or image weighting factors) ensures the tissue content is the same at each of their corresponding voxels. Without this step, the benefits of multi-valued data – robustness to noise and improved separability of tissue classes in the intensity space (contrast) – are eliminated. The registration process (Fig 1.17) is an optimisation task performed over a chosen transformation space, and it usually starts with low-parameter rigid-body (including translation, rotation and uniform scaling) or affine (also including independent axis scaling and skew) transformations to correct global differences. A non-rigid local warping step may follow with a large number of control points to optimise; typically this step is not performed when registering structural volumes due to the low expected amount of non-rigid motion, the high computational cost, and the suitability of the transformations can be poor.

Transformations are made possible through interpolation techniques, which estimate the intensities at new data sampling points: an interpolation kernel or functional form can be assumed between known data points (such as linear, cubic, windowed sinc or spline) to obtain different performance in the spatial and frequency domains. In this

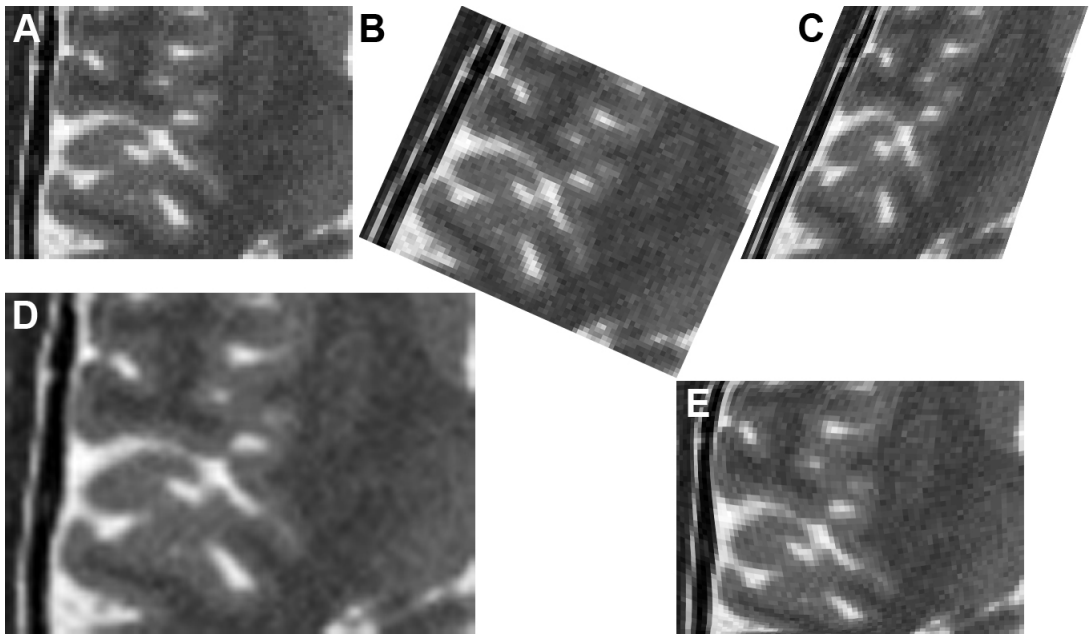

Figure 1.17: **Transformation types**. Transforms are shown applied to a region of an axial slice of the T_2 BrainWeb [Cocosco et al., 1997] phantom displayed in **A**. **B**) Rigid body (rotation). **C**) Affine (skew). **D**) Upscaling with bicubic interpolation. **E**) Non-rigid deformation.

project nearest neighbour and cubic interpolation methods have been used to cause an increase in resolution (upscaling). The reduction of resolution is usually called down-sampling and can be implemented very simply by using an averaging filter.

To obtain the best co-registration a measure of similarity between the volumes must be calculated. Intensity and gradient features are commonly used, but there is also a variety of intensity similarity metrics available which are robust to changes in contrast between the modalities or image weighting factors, including correlation-based (i.e. [Roche et al., 1998]) and information-theoretic (such as mutual information [Studholme et al., 1999][Crum et al., 2003]) methods. Labelling-based similarity metrics allow for direct comparisons of the estimated tissue contents rather than checking information-theoretic or statistical correspondence [Zitova and Flusser, 2003]. Naturally, the intensity, shape and locality models of segmentation – which can make the labelling robust to imaging noise and artifacts – determine the accuracy of this type of metric but many single MRI weighting factors provide poor contrast between some tissue types, thus rendering this method non-viable in some regions.

1.3.3 Pre-processing pipeline

Following conversion of the data which is output from the MRI scanner to appropriate file or data formats for processing, the preparation of MRI dataset (perhaps with several image weighting factors) for segmentation involves three main steps. Firstly, bias field correction aims to reduce – or ideally eliminate – slowly varying contrast changes (Fig 1.5). From a signal processing perspective then the high frequency components (anatomical structure) of the true image are reduced and low frequency spatial variation is increased by bias, and from a probabilistic modelling view the intensity histogram becomes less sharp and the entropy increases. The approaches can be broken down into strictly pre-processing methods, which fit smooth multiplicative fields to the volume intensities in order to reduce entropy [Likar et al., 2001] or to eliminate gradual variation in pure tissue regions [Dawant et al., 1993], and those that are coupled with a segmentation algorithm in order to account for the residual difference between a reconstructed image and the original intensities [Wells et al., 1996].

Secondly, the volumes may need to undergo co-registration (detailed in Section 1.3.2) or be aligned with a brain atlas to provide tissue content priors. Smoothing the volumes alleviates the potential to introduce transformations which are nonsensical through matching noise or unique anatomical features – commonly small and thin structures such as cortical sulci and blood vessels. Most often isotropic Gaussian blurring (Fig 1.18B) is used which obeys scale space axioms such as non-creation and non-enhancement of local extrema, but adaptive smoothing such as median filtering (Fig 1.18C) or anisotropic diffusion (Fig 1.18D) may be preferred to reduce noise while preserving some of the detail. Re-sampling of the volumes may be required at this stage so the voxel dimensions of the volumes being registered are made identical: this process is typically known as re-slicing since the slice thickness is often greater than the square in-slice voxel dimensions.

Finally, skull stripping or brain extraction removes all data outside the brain parenchyma (Fig 1.19) that does not bear any useful information. As well as increasing the robustness of subsequent processing by ensuring only parenchymal tissues remain, the number of valid voxels and in turn the processing times can be considerably reduced. The most common automated approach, exemplified by the Brain Extraction Tool (BET) [Smith, 2002], is to model the brain as a single mesh object and to expand its boundary to the skull which forms a dark boundary. Tests on T_1 volumes [Lee et al., 2003] have showed that the results of more time-consuming and interactive semi-automated brain

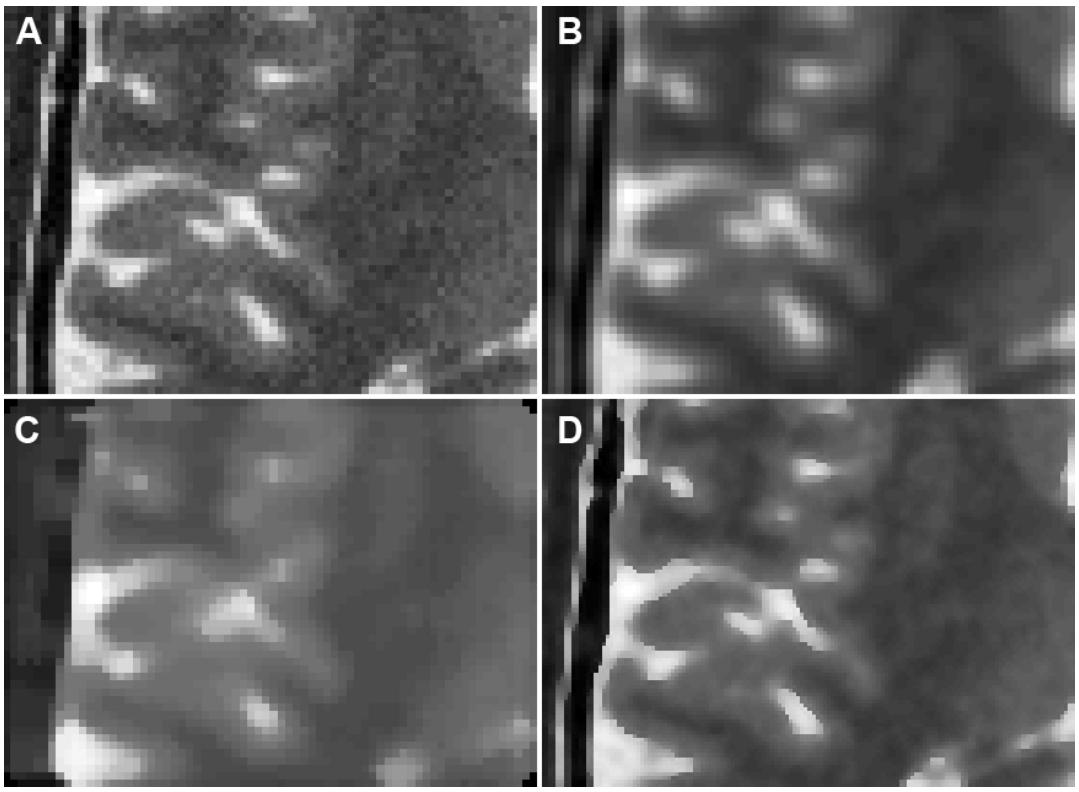

Figure 1.18: **Volume smoothing**. **A**) Unfiltered region from a T_2 axial slice of the BrainWeb [Cocosco et al., 1997] phantom. **B**) Isotropic Gaussian smoothing of **A**, $\sigma = 1$. **C**) Median filtering of **A**. **D**) Upscaling and anisotropic diffusion smoothing of **A** using the method described in [Salvado and Wilson, 2006] with default parameters.

extraction methods were superior to those of the automated ones, including BET and the edge-detection-based BSE [Shattuck et al., 2001], but the differences could usually be corrected without manual editing by adjusting their parameters.

1.3.4 Acquisition of ground truth labelling

Manual segmentation of clinical MRI brain volumes by experts can prove slow and inconsistent by both inter- and intra-rater metrics [Aboutanos et al., 1999], but it is the most practical alternative to *ex vivo* processing (vivisection, fixation and histological staining) which is not viable for human clinical brain imaging. Furthermore, deformation of the brain tissues may occur between imaging and staining which means a non-rigid registration step must be applied; hence the labelling may not be precisely representative of the imaged volume. It is extremely difficult for phantoms (Section 1.1.6) to have the same complexity and same material characteristics as a brain and models of models of noise, the bias field and pathology may not be of sufficient real-

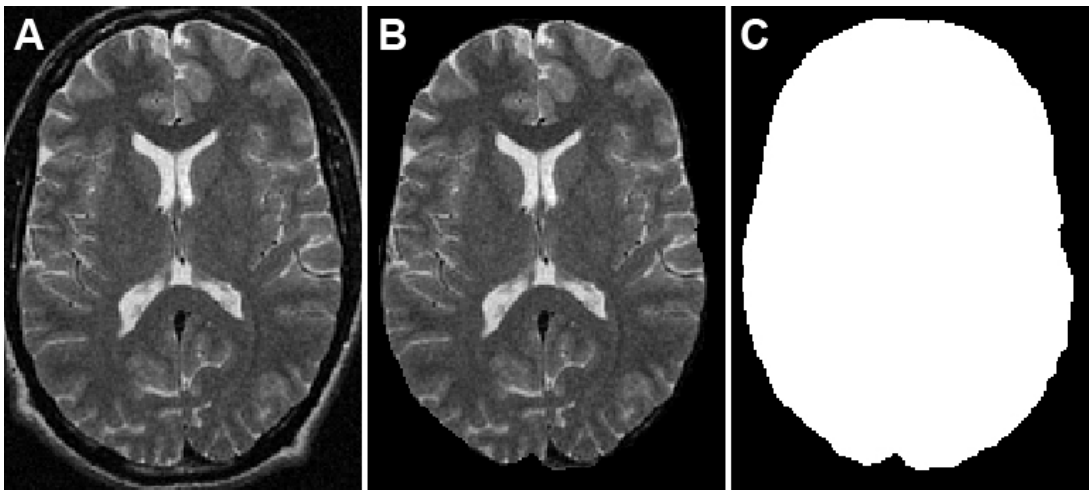

Figure 1.19: **Brain extraction**. **A**) Axial slice of a T_2 volume from the LBC1936 dataset. **B**) **A** after brain extraction using BET [Smith, 2002] has been applied. **C**) Brain mask produced by BET to form **B**.

ism, thus limiting their applicability in testing.

Expert labelling produces a ”surrogate of truth” since it is an estimate based on visual cues and informed judgement (examined more closely in Section 2.3). Due to the poor inter- and intra-rater consistency of expert labelling then averages can be taken over repeated results from the same expert or among different experts. The results typically consist of outlines of the borders of key structures (Fig 1.20); subsequently, flood fill operations can then cover the enclosed spaces. However, the segmentation of thin and overlapping structures can prove difficult using this method, and so less commonly, tissue labels can be assigned on a per-voxel basis (”painting” the brain volumes) to resolve this issue.

1.3.5 Issues with current structural MRI brain volume segmentation methods

Two of the most widely used brain volume segmentation tools in the medical imaging community are FSL FAST[10] and SPM Segment [11]. The FSL package runs in a UNIX environment or requires compilation of the C code, while SPM requires the popular

[10]At the time of publication the version reviewed was 4.1, available from FMRIB at the University of Oxford. A publication [Zhang et al., 2001] describes a previous version but the software and the segmentation algorithms have been updated many times since.

[11]At the time of writing the version reviewed was in SPM5, described in [Ashburner and Friston, 2005] and available from the Wellcome Trust Centre for Neuroimaging at University College London. The software has since been superceded by SPM8 but no updated publication has been published.

Figure 1.20: **Manual anatomical outlining**. A manually-drawn region of interest (ROI) covering the putamen on a T_2 axial slice of the BrainWeb [Cocosco et al., 1997] phantom.

mathematical programming and image processing platform MATLAB. In order to encourage usage and further development, the code developed for this project should also be made available in an accessible format.

Most structural MRI brain segmentation methods in the literature require T_1 data (though FAST and Segment can process other image weighting factors), due to its excellent contrast between GM and WM. These tissues constitute the majority of the parenchymal volume and are also often of most interest to clinicians. Since blood (BG) and CSF both appear dark in T_1 it is difficult to distinguish between sulci, blood vessels, air, the skull, and certain pathologies such as oedema and haemorrhage. In this project dual-echo T_2-PD data will be employed which can provide contrast between the CSF and BG classes, while maintaining some WM/GM contrast.

FAST incorporates neighbourhood context through a MRF prior optimised by ICM, but while FAST can employ tissue content prior maps to improve the accuracy of its initialisation, Segment is reliant upon them to provide robustness to noise. A non-rigid registration process is used by Segment to attain a close-fitting alignment. More finely-grained and purely data-derived label estimation renders FAST a superior tool for labelling detail in structural volumes, whereas SPM is preferred for functional imaging where contrast and resolution can be poor. The results from FAST will thus be the benchmark in this thesis, and a specific framework [Udupa et al., 2006] for describing and comparing computer-aided visualisation and analysis algorithms will be followed to measure the performance of the algorithms studied and to describe their methodologies. The key metrics identified in this framework were:

- The **precision** of forming results – for automated algorithms the processing must necessarily not contain stochastic elements so that results can be exactly reproduced;
- The **efficiency** in clinical settings – expressed in terms of computational complexity and running times on ROIs or volumes;
- The **robustness** of the algorithms – ensuring that an acceptable result was produced given a range of acquisition and labelling conditions, or abnormal (i.e. pathological) data;
- The **accuracy**, involving testing on real-world data which are representative of the problem domain, where ideally a ground truth is supplied in order to allow quantitative evaluation.

10 ROIs from the LBC1936 dataset (described in Section 2.1.1) were thus segmented using FAST to assess its shortcomings in light of the expert-supplied ground truth. This dataset provided a real challenge since the brain volumes possessed poor contrast in some regions as well as age-related degeneration. The labelling and acquisition protocols for the volumes and setting of parameter values for FAST are detailed in Sections 2.1 and 2.5. FAST requires `Hyper` and `mixel` parameters which are not particularly well explained in the documentation; ideally the effect of parameters should be clear and their meaning intuitive, and a framework should be made available for their tuning. Proper selection of these parameters controlling the strength of the MRF spatial interactions must strike a balance between revealing structural detail with weak interactions, and suppressing noise and low frequency variation due to bias with strong values, as well as ensuring mixtures on tissue boundaries are correctly disambiguated.

A primary concern was that the segmentation was robust. There were no degenerate results (where the memberships of one or two classes were dominant and the accuracy of the labelling was extremely poor) for any ROI. Accuracy was also of crucial importance, and examining the absolute error in class fractional contents against an expert's labelling over the ROIs (Fig 1.21), it was found that most of the error occurred due to:

- Widespread misplacement of structure boundaries (marked with blue arrows), especially between white and gray matter;
- Insensitivity to thin structures including blood vessels, the falx cerebri, and cortical sulci (green arrows);
- Insensitivity to sub-cortical structures (yellow arrows).

There may be sub-optimal contrast at WM/GM boundaries in sub-cortical and posterior regions (ROI 2), but due to the poor results that were produced then the viability of both the Gaussian probability distribution of the class intensities and the linear mixing models must be scrutinised. Improving the quality of the data – especially the consistency of strong region boundaries between the T_2- and PD-weighted data – and thus simplifying segmentation decisions to be made may improve the accuracy of the result. Hence, pre-processing methods to achieve volume smoothing in homogeneous tissue regions, PVE reduction at region boundaries, and the proper resolution of contrasting thin structures will be investigated.

The last two findings stated above suggest that the partial volume estimation model employed by FAST may not perform optimally for structures composed entirely of

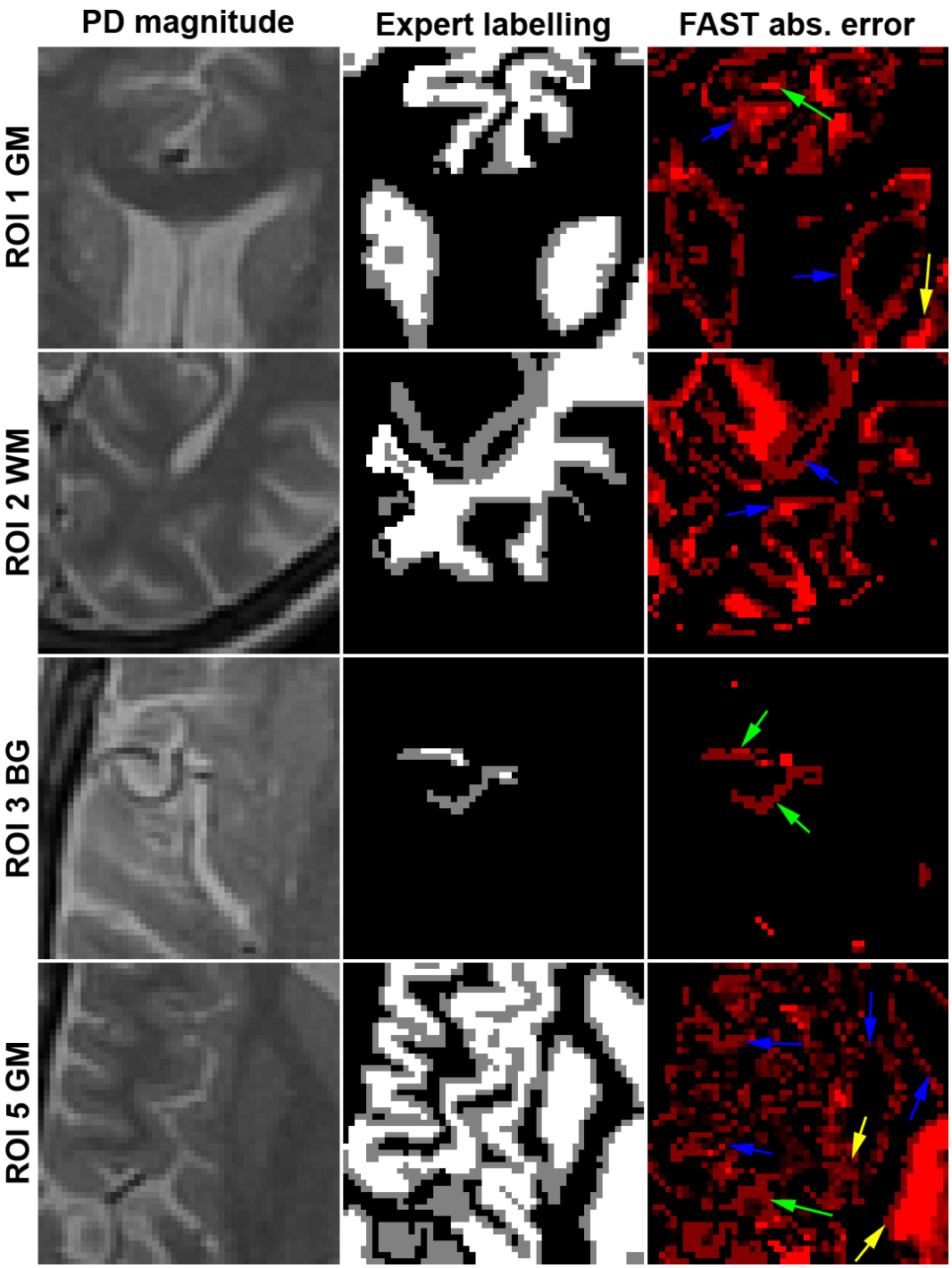

Figure 1.21: **Absolute errors of FAST [Zhang et al., 2001] over LBC1936 dataset ROIs**. **Left**: PD data. **Middle**: Expert labelling for the tissue class mentioned in the row header. White indicates pure content, black indicates none, and gray indicates partial volume. **Right**: Shades of red indicate the absolute error between FAST and the expert. Bright red is a total mismatch, whereas black denotes identical labelling. Blue arrows: misplacement of structure boundaries. Green arrows: insensitivity to thin structures. Yellow arrows: insensitivity to sub-cortical and narrow structures.

partial volume voxels. In regions of pure tissue the uniformity of the labelling was generally very good due to the MRF prior, at the expense of sensitivity to small oedemic structures (CSF-filled blobs observed in ROI 1) and thin structures. This suggests the potential for improvement in the label estimation and the neighbourhood interaction procedures via incorporation of class content priors or specialised filtering for small or thin structures. Since most brain atlases cannot capture the unique topology of such brain anatomy (without deformable and adaptable aspects incorporated within their form, as discussed in Section 6.4.1) then the priors will need to be derived from the data itself. These priors can also help control against the low population of voxels with BG or CSF content and thus improve robustness. There is also potential for coupling the thin structure identification process to the data quality enhancement method, since it is important that the thin structures are preserved (or their resolution even restored) during volume smoothing.

FAST completes its processing of dual-echo structural brain volumes in relatively short running times, ranging from 15 to 30 minutes with no other programs running[12] depending on the number of slices in the volume and the number of iterations performed. New methods which are developed in this thesis should also endeavour to be able to run on a single desktop computer and to complete the segmentation of a full brain volume in the order of hours (rather than days), which will make future development and testing a manageable task. Since the objective functions of more complex segmentation methods could be extremely non-convex – which makes the optimisation time-consuming and prone to finding poor local minima – good initialisation is essential, but still the optimisation may have to employ relatively simple methods.

Another issue stemming from hardware constraints concerns memory limitations: since clinical brain volumes typically contain around 2^{23} ($256 \times 256 \times 128$) voxels, and modern desktop computers possess around 2^{32} bytes of RAM, then up to 512 bytes of information can be stored for each voxel and accessed quickly. In order to run efficiently on a desktop computer then the amount of memory used should be tightly controlled, especially if volumes are resized or tensors are produced. However, the 64-bit operating systems used to develop this work also allow the hard disk space to be addressed, providing more than 2^{40} bytes of further storage, albeit with much slower access.

[12]The specifications of the computer on which this processing was performed are listed in Section 2.6.4.

1.4 Aims, objectives, structure and contributions of this thesis

The evaluation of FAST in Section 1.3.5 gave rise to one primary aim for a better dual-echo MRI structural brain volume segmentation algorithm: performance around thin and sub-cortical structures should be improved. The outstanding areas of research identified that could fulfil this aim were the specific identification of thin cortical sulci and blood vessels in the data, and the enhancement of the data quality in order to improve the contrast of both thin and sub-cortical structures as well as the sharpness of their boundaries. These methods could be incorporated as enhancements to the ideas expressed in the unified segmentation method [van Leemput et al., 2003] (examined in Chapters 4 and 5) which did operate at a sub-voxel scale and could enable PVE reduction to simplify the segmentation process, but it was not able to properly express the labelling of thin or partial volume structures; hence, other improvements to allow the proper labelling as well as optimise the labelling of small-scale structures correctly were required. The four evaluation metrics identified in Section 1.3.5 (precision, efficiency, robustness and accuracy) also contribute to the formation of the thesis objectives listed below.

- The algorithms should be designed to process dual-echo T_2-PD MRI structural brain volumes, which can provide good BG/CSF contrast for thin cortical sulci and blood vessels;
- Data-derived tissue content priors should be created to help ensure the robustness of segmentation and also improve accuracy around cortical sulci and blood vessels, since atlas-based priors cannot currently be made sufficiently relevant for these thin structures which have unique configurations between individuals;
- Adaptive smoothing and PVE reduction of the volume data – using the tissue content priors to aid intensity restoration processes – should be performed to enhance the quality of the data and make it more amenable to accurate segmentation;
- The neighbourhood interaction and label optimisation models should allow rapid variation across region borders and thin structures but prefer homogeneity in pure tissue regions, which may require adaptive and deliberate processing when considering thin structures;

- Precision of the results is an important requirement so formal training procedures should be introduced for important parameters, and non-stochastic methods should be employed to ensure repeatability of generating results;
- Testing should involve quantitative evaluations performed on a wide range of data (from which robustness can be inferred), including expert-labelled MRI volumes acquired from a scanner as well as pathological and aged brain tissue.

Some less essential goals are also presented below.

- The viability of Gaussian probability distributions for pure tissue intensities and the mixel model should be investigated over the given datasets;
- The efficiency should be recorded, in terms of running time on a single modern desktop computer as well as rough estimates of computational complexity;
- Processing should be able to be performed on a single desktop computer, with the total time spent on a single volume in the order of hours rather than days;
- The memory constraints of modern desktop computers should be appreciated;
- The parameters required should be meaningful and documented well.

This thesis consists of six chapters and an Appendix containing supplementary figures and tables. The methods chapters (Chapters 3-5) possess demonstrations and discussion of the performance of their proposed methods (also covering the last four minor objectives), both in isolation and in the context of the other chapters' work.

- The **data acquisition and method evaluation strategies** chapter tackles the final main objective, to detail the data and labelling acquisition protocols for all the datasets processed, and to allow meaningful evaluations of results in the following chapters. A range of quantitative error measures and qualitative evaluation guidelines are stated for assessment of the accuracy of the methods. The design of a new bespoke software tool to assist production of the ground truth labelling by a human expert is also presented.
- The **thin contrasting structure detection and anisotropy-preserving filtering** chapter tackles the robust identification of thin cortical sulci and blood vessels from dual-echo T_2-PD MRI structural brain volumes (fulfilling the first two main objectives). A particularly important improvement to multi-scale vessel

enhancement filters [Frangi et al., 1998] lies in the adaptation of tensor regularisation methods [Westin and Knutsson, 2003] to smooth the maps and expand their coverage. Orientation and scale preferences to preserve the thin structures in future smoothing and neighbourhood interaction processes are implemented as automatically-produced anisotropic smoothing filters. Procedures are outlined to remove large contrasting structures from the maps as well as to train the range of thin structure scales investigated (dealing with the second-to-last main objective).

- The **thin-structure-preserving volume resolution enhancement** chapter details several developments applied to the methods recently proposed in [Salvado et al., 2006][Salvado and Wilson, 2006] for adaptively smoothing MRI volumes and reducing PVE (fulfilling the third main objective). The modifications include dual-channel processing, the restoration of thin structure intensities, and the application of the anisotropy-preserving smoothing filters produced in Chapter 3 (first and fourth main objectives). The context of this work is discussed in relation to a unified segmentation method [van Leemput et al., 2003], used as inspiration for the work in Chapter 5.

- The **thin-structure-preserving soft segmentation of brain tissues** chapter describes the automated segmentation method for the resolution-enhanced dual-echo T_2-PD MRI structural brain volumes. It uses a deliberate relaxation labelling neighbourhood optimisation scheme [Li et al., 1997] in an EM-based framework for soft segmentation [Liang et al., 2003][Liang et al., 2007] to improve segmentation accuracy around cortical sulci and blood vessels in particular, as well as sub-cortical structures (fulfilling the primary aim as well as the first and fourth main objectives). The BG and CSF thin structure maps produced in Chapter 3 form data-derived tissue priors, and the anisotropy-preserving filters play a role in the neighbourhood label optimisation of thin structures. In this chapter the viability of using the mixel model and Gaussian probability distributions for pure tissue intensities on the LBC1936 dataset was briefly investigated (fulfilling the first minor objective). The testing is the most extensive in this chapter, and examines the performance and robustness of the complete system on both the aged LBC1936 cohort (also including a severely atrophied brain) and the BrainWeb [Cocosco et al., 1997] phantom (both the regular and multiple-sclerosis-lesioned variants with varying levels of noise). Comparisons

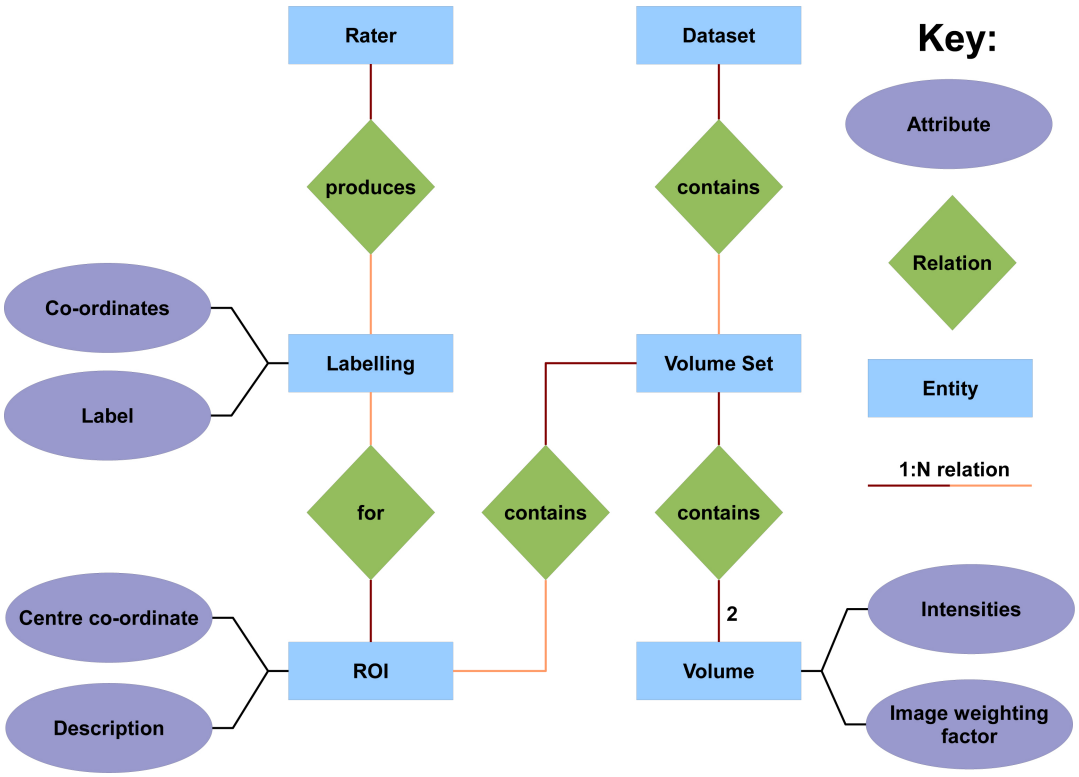

Figure 1.22: **Entity relationship diagram for data in the soft-labelled MRI brain volume corpus**. Subsystems are highlighted in light red.

of performance are made against the FAST [Zhang et al., 2001] segmentation tool and the unified segmentation method [van Leemput et al., 2003].

- The **conclusions** chapter contains a final summary and evaluation of the results presented throughout the thesis with respect to the objectives listed in this section, as well as more detailed proposals for improvements or alternative applications for the methods beyond those discussed in each chapter. These topics include the potential of triple-channel processing, the formation of data-dependent contour-based brain tissue priors and atlases, modifications to the STAPLE [Warfield et al., 2004] algorithm to allow consensus formation from soft segmentation results, and the utility of soft labelled data as a similarity metric for single-subject multi-modal volume co-registration. Code which implements the algorithms presented in this thesis will be made freely available on Sourceforge as part of the *RESCUE segmentation* project, with more details presented in this chapter.

The main data entity relationships are presented in Fig 1.22. Automated processes or human experts can label a volume, where each voxel which is both within the brain

mask and inside a ROI is assigned a soft label. A volume set contains a pair of T_2- and PD-weighted volumes from the same subject, and a dataset consists of one or more volume sets.

A system flowchart, describing data flow and operations conducted in the methods which are described in this thesis in order to solve the issues identified in Section 1.4, is presented in Fig 1.23. The dual-channel data is acquired to provide volumes with contrast for both BG and CSF tissues (in the PD- and T_2-weighted data, respectively) as input for the thin structure detection and volume resolution enhancement processes. BG and CSF tissue prior maps as well as shaped and oriented filters for thin structures are extracted and are used in the subsequent volume resolution enhancement process to preserve and even restore the thin structures' intensities. At this stage adaptive smoothing, PVE reduction and contrast enhancement is performed on both data channels simultaneously to ensure a consistent result that should simplify the segmentation process. Finally, the output of both previous stages enters the main segmentation algorithm. Here, the tissue priors help to inform the labelling and the relaxation labelling step (R-step) – as part of the expectation maximisation framework (E- and M-steps) – should allow rapid variation in the labelling to occur around region borders and thin structures, but prefer homogeneity in pure tissue regions.

1.5 Summary

In this chapter several background topics have been covered in Sections 1.1-1.3 as necessary prerequisite knowledge for the understanding of future chapters. While each chapter has its own introduction and summary of the relevant literature, detail on the basics of MRI acquisition and brain volume segmentation as well as an overview of brain anatomy are presented here so that researchers in the field may find the future chapters more compact and relevant.

Supplementing these introductory topics are the formation of aims and objectives for the thesis in Section 1.4 as a result of the analysis of FAST on the LBC1936 data in Section 1.3.5. Finally, the structure of the thesis and the segmentation framework are presented in Section 1.4 for ease of reference.

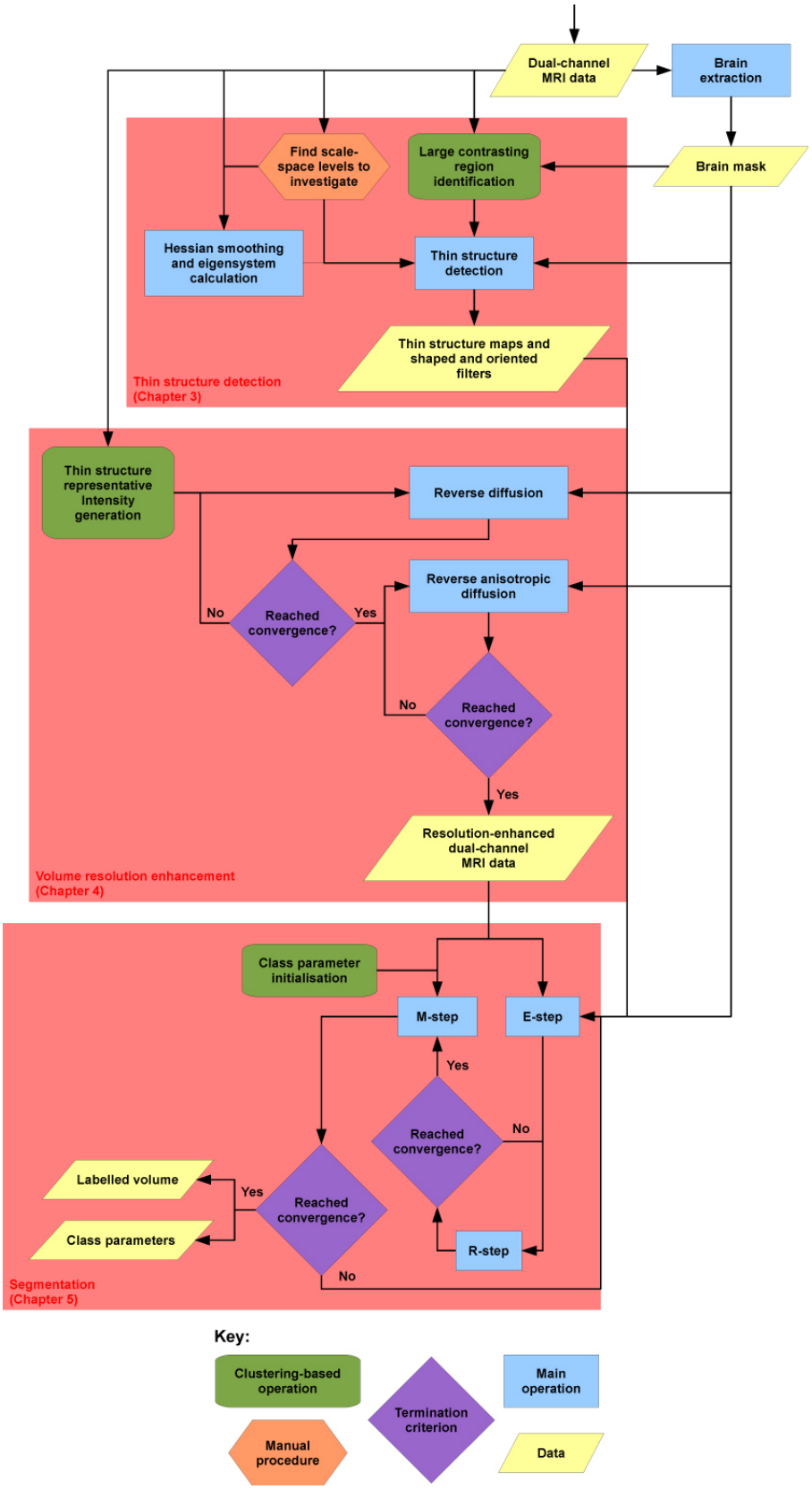

Figure 1.23: **Systems flowchart for the methods proposed in this thesis**.

Chapter 2

Data acquisition and method evaluation strategies

To avoid repetition in the following three methods chapters then all the protocols for data acquisition and pre-processing, the production of the ground truth labelling for the LBC1936 dataset, and quantitative as well as qualitative evaluation of the results are listed in this chapter. These topics provide the means of assessing the key metrics identified in Section 1.3.5: precision, efficiency, robustness and accuracy. Dr Andrew Farrall, a consultant radiologist at the Western General Hospital[1] was instrumental in providing requirements and feedback for a software tool designed to assist his labelling task, and also for performing manual labelling on a MRI dataset of aged brains acquired in clinical settings.

2.1 Acquisition of test data

Physical phantoms and simple software phantoms – containing polygons or other basic filled shapes, examined in Section 1.1.6 – have not employed since they cannot be sufficiently representative of real brain anatomy. Instead, the BrainWeb [Cocosco et al., 1997] software phantom (also described in Section 1.1.6) was used to investigate performance under ideal acquisition and contrast conditions. No intensity non-uniformity effects were applied so bias correction – which does not test the segmentation algorithm – was not needed. T_2 and PD volumes with several different additive noise levels (1%, 3%, 5% and 9% of the maximum volume intensity) and 1mm^3 cubic voxels were produced, and the areas surrounding two ROIs located on slices 65 and 75 were

[1] Edinburgh, UK.

particularly closely investigated. Furthermore, $2 \times 2 \times 2$ downsampled volumes (produced using a similar method to that expressed in Eqn 5.5) were processed to assess the method's performance both on low-resolution data and also on entire volumes. A version of the BrainWeb phantom has also been made available with additional multiple sclerosis lesions, and similar parameters were used to generate volumes for testing the methods' robustness to pathological tissue. In these volumes, two ROIs located on slices 75 and 105 were investigated.

Testing on brain MRI volumes acquired in clinical settings has been advocated in the frameworks described in [Udupa et al., 2006] to ensure real-world viability of methods. Since no public repository of soft-labelled dual-echo T_2-PD MRI brain volumes exists[2] then data from five randomly-selected subjects in the Lothian Birth Cohort (LBC1936), a MRI dataset of brain images acquired from male subjects aged 71-72 at the Western General Hospital[3], was accessed. This data (studied more closely in Section 2.1.1) incorporated age-related brain degeneration as well as poor contrast in many regions. One subject in LBC1936 with extensive brain atrophy was also selected to provide a pathological dataset.

The scanner used to acquire the LBC1936 dataset was a GE Signa LX 1.5 Tesla clinical scanner present at the Western General Hospital[4]. A birdcage head coil was used with padding to prevent as much head movement of the subjects as possible. The examinations included a dual-echo Fast Spin Echo sequence for the acquisition of dual-echo T_2-PD structural volumes. The voxel size was 1mm $\times$ 1mm with a 2mm slice thickness, there was no inter-slice gap, and other scan parameters were NEX = 1, TE_{PD} = 9ms, TE_{T2} = 102ms, TR = 11320ms, FOV = 256mm and matrix size 256 $\times$ 256.

Two further datasets of my own brain were acquired so the scanning conditions could be simplified: no time constraints had to be enforced, the amount of sample motion would be very low, and young healthy brain tissue would be present. The same scanner and head coil were used as for the collection of the LBC1936 dataset, and both datasets were acquired without inter-slice gaps. In comparison to the LBC1936 dataset (Fig 2.1) tissue-specific peaks in the intensity histogram were better defined, thin structures appeared sharper (marked with blue arrows), and particularly in occip-

[2]However, hard labelling of T_1 data is more common, such as the Internet Brain Segmentation Repository provided by the Center for Morphometric Analysis, Massachusetts General Hospital and available at `http://www.cma.mgh.harvard.edu/ibsr/`.

[3]Edinburgh, UK.

[4]Edinburgh, UK.

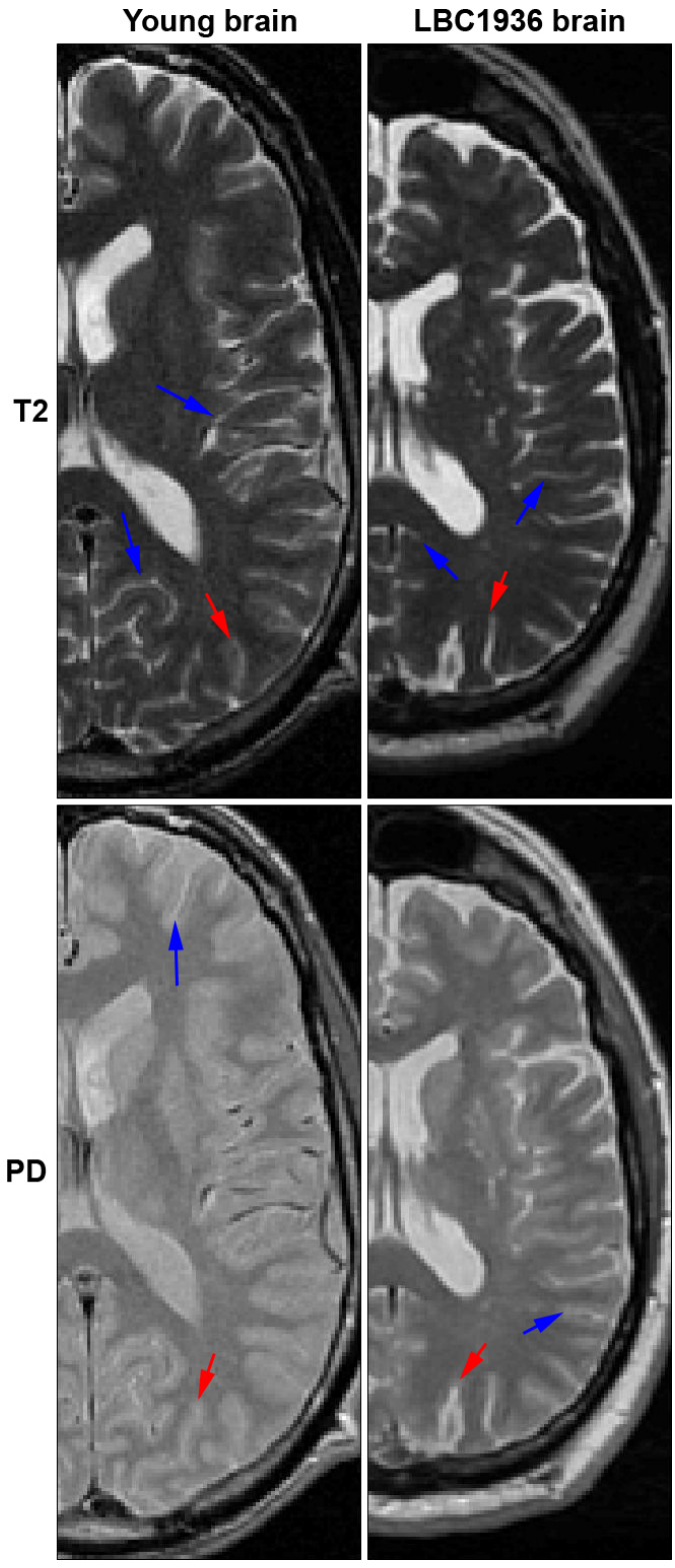

Figure 2.1: **Comparison of LBC1936 and young brain dataset quality**. **Left column**: LBC1936 data. **Right column**: Young brain data. **Top row**: T_2 data. **Bottom row**: PD data. Coloured arrows are referenced in Section 2.1.

ital regions (red arrows) WM/GM contrast was improved – though GM/CSF contrast was not as good in PD.

The first dataset consisted of a dual-echo T_2-PD FSE sequence with a 1mm $\times$ 1mm voxel size with a 2mm slice thickness, NEX = 1, TE_{PD} = 15ms, TE_{T2} = 105ms, TR = 3600ms, FOV = 256mm, and matrix size 256 $\times$ 256. The second was acquired at two different resolutions: the first acquisition had a 256 $\times$ 256 matrix and a 2mm slice thickness, and then subsequently a 512 $\times$ 512 matrix and a 1mm slice thickness were used to double the resolution. To combat the reduction in SNR of the high resolution data, this acquisition was formed from the average of multiple volumes (NEX = 7). The co-registration process for the multiple excitations is detailed in Section 2.2 and the other parameters common to these dual-echo T_2-PD FSE sequences were TE_{PD} = 20ms, TE_{T2} = 110ms, TR = 3240ms and FOV = 256mm.

2.1.1 Selection of ROIs from real brain volumes

A major consideration in using human experts to assign soft labels to MRI data is the huge number of data points ($\approx 2^{23}$ or 8 million in a $256 \times 256 \times 128$ volume) to be individually processed. Even if single voxels could be labelled in a few seconds each, labelling the entire brain is unrealistic: a conservative estimate that 30% of voxels are removed by skull stripping means that over 5 million will still remain to be processed for a single volume. To reduce the workload, it was agreed with Dr Farrall that the labelling would be restricted to ten 50 $\times$ 50 voxel ROIs, with two ROIs defined on each of five brain volumes selected at random from the LBC1936 cohort. The ROIs (which have been displayed and their anatomy labelled in Figs 2.2 and 2.3) were defined on single axial slices and were chosen to contain structures that commonly pose problems to segmentation methods:

- **Corpus callosum** – ROI 1 (Volume 1), and ROIs 7 and 8 (Volume 4). The most dense concentration of WM axons in the brain should remain uncorrupted by noise, and partial volumes at the interfaces with GM and CSF should be correctly quantified;

- **Falx cerebri** – ROIs 1 and 2 (Volume 1), ROI 7 (Volume 4), and ROIs 9 and 10 (Volume 5). This region separating the hemispheres contains thin sulci, blood vessels, cortex and WM;

- **Lateral ventricle** – ROIs 1 and 2 (Volume 1), ROI 6 (Volume 3), ROIs 7 and

8 (Volume 4), and ROI 9 (Volume 5). This is the largest body of pure CSF and mixing with WM and GM exists at its boundaries. Within the ventricle is the choroid plexus, a structure with BG-like intensity;

- **Temporal cortex** and **insula** – ROI 3 (Volume 2) and ROI 5 (Volume 3). The cortex in this region is particularly convoluted, creating complicated GM/CSF and WM/GM boundaries. Blood vessels are pervasive, thin regions of WM are present, and there are a range of sizes of sulci, from the very small to the lateral sulcus;

- **Occipital cortex** – ROI 2 (Volume 1) and ROI 10 (Volume 5). This region of cortex is thinner than in frontal areas [Fischl and Dale, 2000] and the WM/GM contrast tends to be much poorer, complicating accurate segmentation of these classes.

- **Deep gray matter regions** – ROI 3 (Volume 2), ROIs 5 and 6 (Volume 3), and ROIs 7 and 8 (Volume 4). These nuclear structures aggregate within the WM near the ventricles and appear to possess intensities indicative of WM/GM partial volumes. They exhibit a range of sizes and shapes, from large ellipsoids (thalamus) to thin lamina (claustrum) and small blobs (globus pallidus);

- **Cerebellum** – ROI 4 (Volume 2). Its extremely convoluted cortex and complicated WM folia mean that the small-scale tissues are poorly resolved in clinical scans. GM/CSF and WM/GM partial volumes will be abundant in this region.

2.2 Pre-processing

MATLAB was chosen as the development platform for the proposed methods, and so MRI data files in the scanner's DICOM format were converted to separate T_2 and PD volume matrices using the MATLAB `dicomread` function. To convert between MATLAB matrices and the NIfTI file format required by the competing software package FSL, two functions `read_nii` and `write_nii` from the input/output package *Tools for NIfTI and ANALYZE image*[5] were used.

Co-registration was needed for the high-resolution young brain dataset, which was acquired with multiple excitations (NEX > 1). The *FSL Linear Image Registration*

[5]Written by Jimmy Shen and available at `http://www.mathworks.co.uk/matlabcentral/fileexchange/8797`.

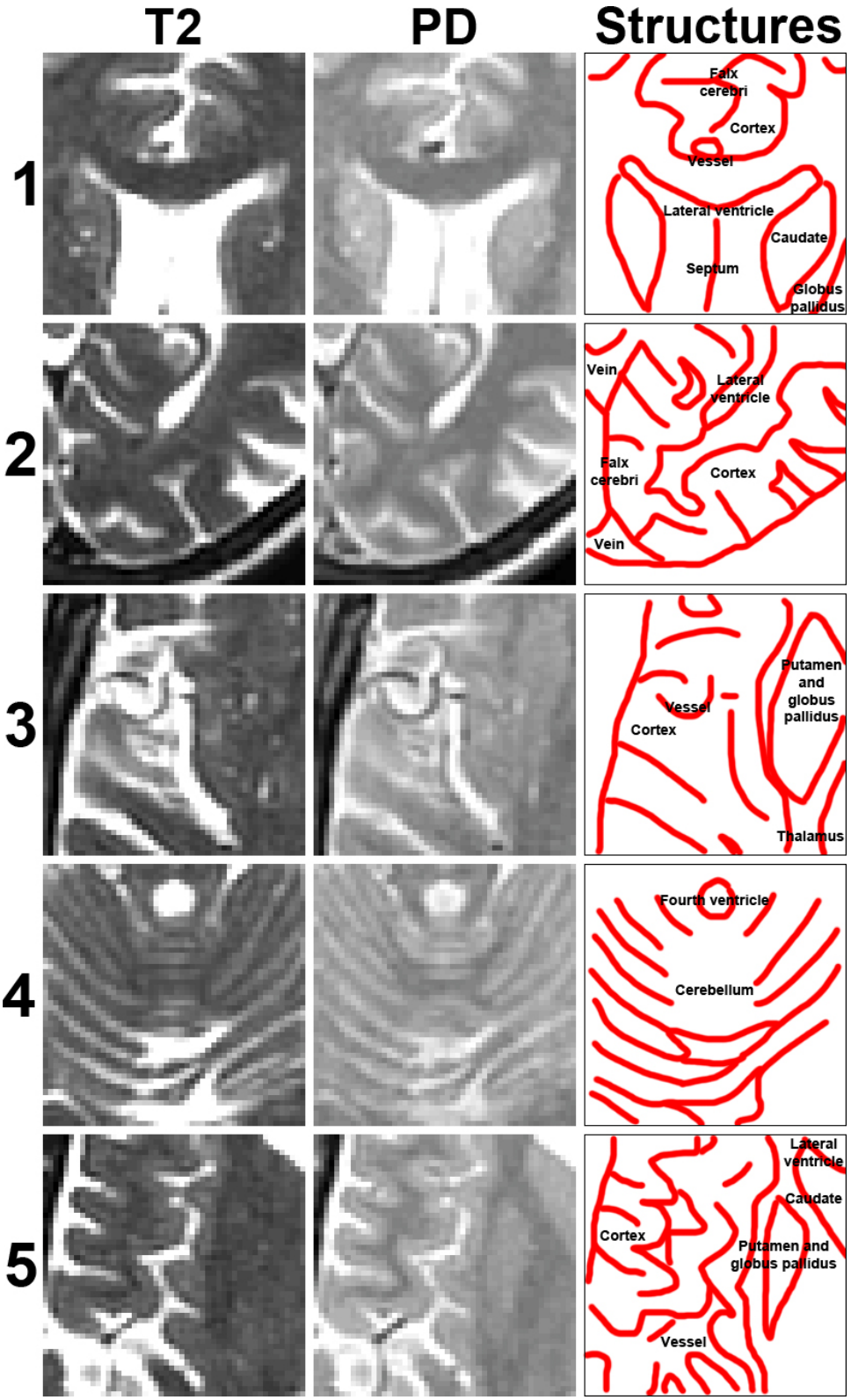

Figure 2.2: **Regions of interest (ROIs) 1-5 chosen from volumes in the LBC1936 dataset**. Each ROI is numbered and structures of interest are labelled in red in the right-most column. WM structures are not labelled.

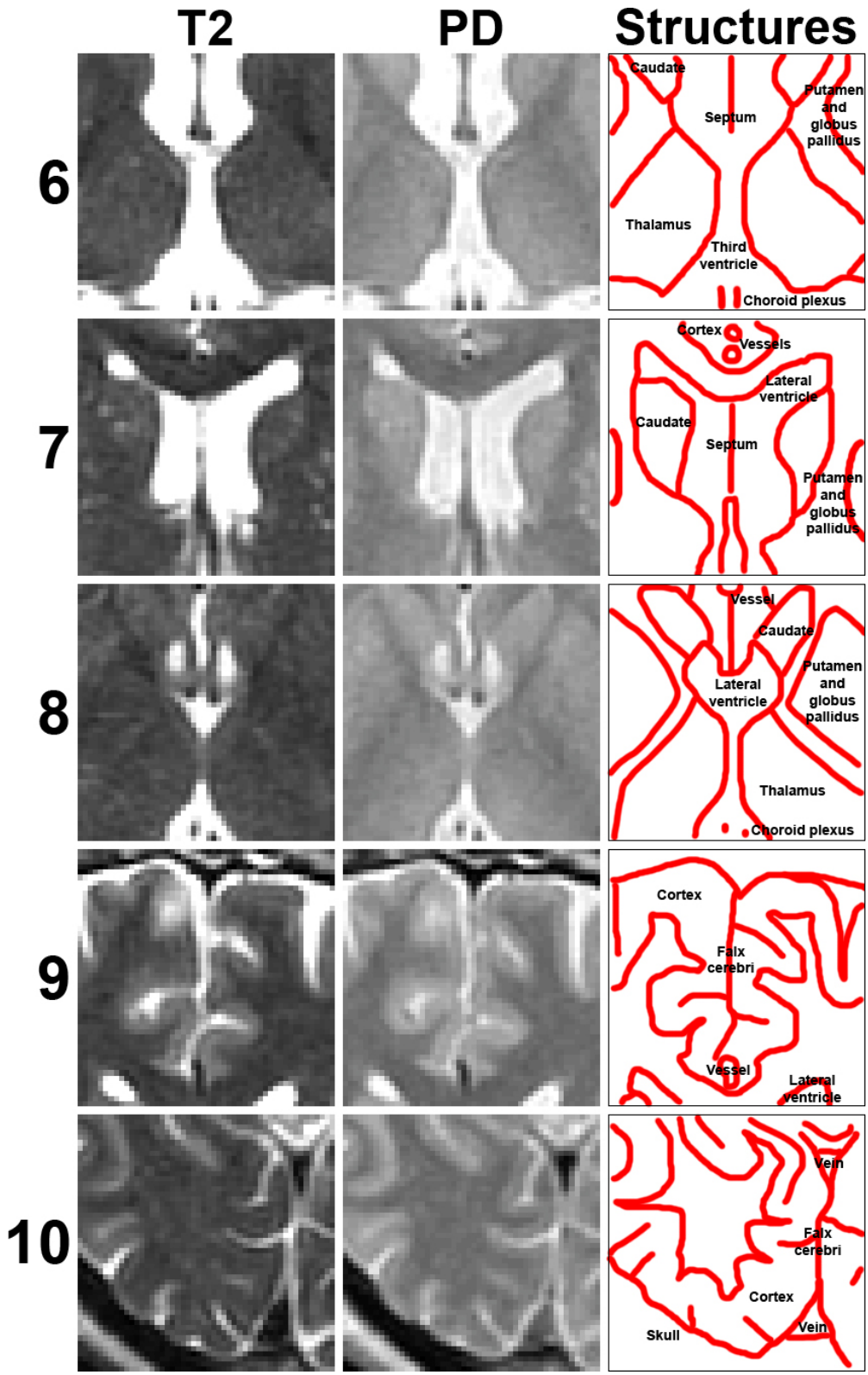

Figure 2.3: **Regions of interest (ROIs) 6-10 chosen from volumes in the LBC1936 dataset**. Each ROI is numbered and structures of interest are labelled in red in the right-most column. WM structures are not labelled.

Tool (FLIRT) [Jenkinson and Smith, 2001] is widely used at the Western General Hospital[6] for correcting rigid registration differences between structural volumes, and it was also used in this work. The parameters passed to FLIRT were `dof` = 12 (assuming only affine transformations had occurred), `cost` = `corratio` and `searchcost` = `corratio` (using the correlation ratio similarity metric with 256 bins), `interp` = `sinc` (sinc interpolation with a Hanning window of 7 voxels), and angular search space limits and steps `searchrx` = -40:40, `searchry` = -40:40, `searchrz` = -40:40, `coarsesearch` = 10 and `finesearch` = 2.

The brain parenchyma was extracted from the volumes acquired from the scanner using the *Brain Extraction Tool* (BET) [Smith, 2002], which is also part of the FSL package and again in widespread use at the Western General Hospital[7]. The fractional intensity threshold `f` was set to 0.5 for PD and to 0.35 for T_2; the tuning of this parameter was a simple task as the skull is fairly easy to recognise as a smooth, well-formed dark region at the edge of the brain. In a short evaluation over ten volumes from the LBC1936 dataset BET was found to work consistently and accurately for these parameter values (data not shown). The conjunction of the brain masks for each T_2-PD volume pair was subsequently morphologically filtered to close small holes and produce a boolean brain parenchyma mask.

Bias correction of these volumes was attempted using BFC [Shattuck et al., 2001] and N3 [Sled et al., 1998], since in a round-up of such methods [Arnold et al., 2001] they were found to give more stable convergence and introduce little additional inhomogeneity. However, the results which were produced were unsatisfactory: increased intensity non-uniformity was found in some regions (particularly WM) despite trying a range of parameter values, and consequently bias correction was not applied. Uniform smoothing to reduce noise was also not performed as a pre-processing step since it would lead to a loss of thin structural details, though anisotropic smoothing was tackled by the proposed volume resolution enhancement method (Chapter 4). As a final pre-processing step, the maximum intensity of each volume was windowed to prevent the top 1% of voxel intensities from forming excessively large outliers.

[6] Edinburgh, UK.
[7] Edinburgh, UK.

2.3 Soft labelling protocols for the LBC1936 dataset

Each voxel in the LBC1936 dataset ROIs required a label to establish a ground truth for future thin structure detection and segmentation methods. A constrained mixel dimension of 2 was chosen, as advocated widely in the literature, and so up to 2 classes described in Section 1.3 (BG, WM, GM and CSF) as well as an *other* class could contribute to a voxel's label. A partial volume of two classes was assigned if any proportion of a second class was judged to be present inside a voxel; otherwise a pure class label was set. The fractional contents were not defined (but assumed to be on average 50%) due to concerns over the quality of the data as well as the accuracy and consistency that could be achieved with more exact quantification given time constraints. Discussion with Dr Farrall led to the formation of the following protocol for labelling small groups of (or individual) voxels:

- All voxels in an ROI should be labelled in a single session if possible, so that a train of thought can be maintained;
- The voxel values in the local 3D neighbourhood must be investigated, in both T_2 and PD;
- The expert's mental models of brain anatomy should be consulted to decide upon the expected tissue types at the location, including assessment of evidence of pathology;
- Optionally, the minimum and maximum intensity windowing should be able to be altered to eliminate gray-level illusions[8] and restrict investigations to surround the estimated intensity mean of the proposed label;
- The estimated constituent tissues then determine which label should be applied to the voxel(s).

2.4 Creation of a software tool to assist soft labelling of volumetric data

Since five possible classes were identified in Section 2.3 then the required labelling format for the LBC1936 ROIs was a binary five-element membership vector at each

[8] An excellent overview on the topic of grayscale illusions can be found in [Adelson, 2000].

voxel. Only one or two classes could be present at each voxel (and consequently only one or two elements of the vector would be non-zero) due to the constrained mixel dimension of two that was selected. For ease of recording and analysing the labelling produced by Dr Farrall, the natural choice was to use a software tool to collect the results from this manual process. With the correct design of the interface for navigating between slices, performing three-plane localisation of neighbouring voxels, and appreciating label choices which have already been assigned then the tool could also help to improve the efficiency and accuracy of labelling.

The presence of very thin, convoluted, or partial volume structures meant that simply outlining anatomical boundaries was insufficient to produce a soft labelling; furthermore, this method did not encourage close examination of individual voxels in a 3D neighbourhood when investigating structure borders and other peculiarities. Instead, the label assignment needed to be performed on a voxel-by-voxel basis in the vein of "paint" software – and therefore the features and user interface of several graphics processing tools (listed below) were investigated in detail for inspiration. Several other tools[9] were also examined but did not offer any features which were sufficiently different to warrant specific mention.

- **Adobe Photoshop**[10] (version CS4), an example of a commercial graphic design package with plug-in support;
- **Analyze**[11] (version 8), a commercial medical image processing suite with which clinical staff at the Western General Hospital[12] are familiar;
- **MIPAV**[13] (version 4.1.0), a free open-source Java-based medical image processing tool with plug-in support;
- **ITK-SNAP** [Yushkevich et al., 2006] (version 1.8.0), an open-source C-based semi-automatic segmentation tool which is part of the Image Toolkit (ITK) platform.

None of the tools studied could be easily adapted to produce an efficient voxel painting interface which could also switch seamlessly between the channels of the T_2-PD data, due to either a lack of access to the source code or because of the significant

[9] Including GIMP, Digital Jacket, ImageJ, 3D Slicer, BrainVisa, MRICron, MeVisLab and MRIVIEW.

[10] Adobe Systems Inc., San Jose, CA, USA.

[11] Mayo Foundation, Rochester, MN, USA.

[12] Edinburgh, UK.

[13] Available at `http://mipav.cit.nih.gov`.

time investment needed to modify existing frameworks. Therefore, a new labelling tool inspired by the aforementioned products was designed in MATLAB. Ease-of-use, efficiency and user feedback were the most important aspects of the design, and since the user was not expected to be especially skilled in working quickly with graphics packages then all input needed to be performed just using the mouse, with keyboard accelerators available once the user acquired more experience.

Minimal complexity in the user interface was key in promoting efficiency, and this aspect of the design was mainly inspired by ITK-SNAP. Its simple and graphically-consistent presentation, combined with the usage of tool interfaces which took a minimum of screen space, contrasted with the cluttered window-based approaches of Analyze and MIPAV. Volume viewers placed prominently on one side of the screen with nearby navigation buttons were able to focus the attention of the user, with tool interfaces gathered on the other side. The viewport had two viewing modes: an *expanded mode* requested by Dr Farrall where only data in the axial plane was shown (Fig 2.4A), and a *three-plane mode* where the axial, coronal and sagittal planes were displayed (Fig 2.4B) similar to MIPAV and Analyze.

The range of interface features available was precisely tailored to the task in order to reduce the time spent locating them. An interview with Dr Farrall revealed that the essential features were coloured overlays of the class content, a three-point (minimum, median and maximum) intensity windowing and contrast control, and a notification of unlabelled voxels. The only other parts of the interface that needed to be readily accessed were a list-based label selection interface and navigation controls, including zooming as well as switching between the different image weighting factors and ROIs. Like Photoshop, rarely-used functions were sensibly relegated to menu options, whereas commonly-used tools made a prominent and consistently-placed appearance (Fig 2.5).

In order to shield the user from dealing with data management (unlike ”sandbox” tools such as MIPAV), all the volume data and ROI locations were pre-loaded and auto-saving was implemented. As thin structures cannot be accurately and easily labelled in 3D using multi-part polygonal regions, spline definition tools, free-form perimeter tracing, graph theory-based livewire [Falcao et al., 1998] algorithms or active contour evolution methods – available in MIPAV, Photoshop and ITK-SNAP – then label assignments for individual voxels, in the vein of the pencil tool in Photoshop, were necessary. Different brush sizes were made available in either single, 3×3 or 5×5 voxel sizes to increase labelling speed. Left-clicking on a voxel in a viewport set the

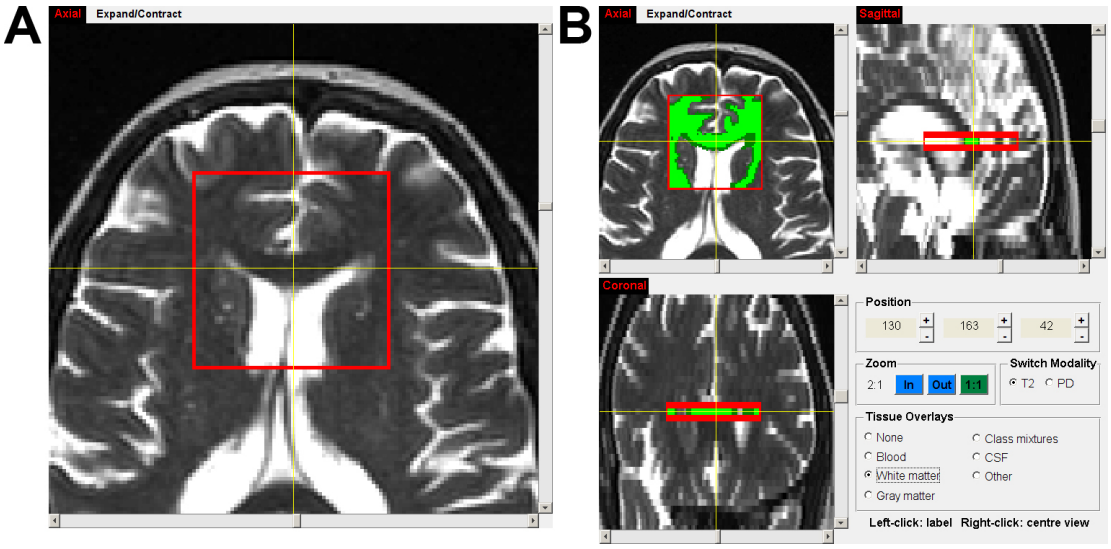

Figure 2.4: **Labelling tool viewports**. **A**) Expanded mode. **B**) Three plane mode with WM overlay.

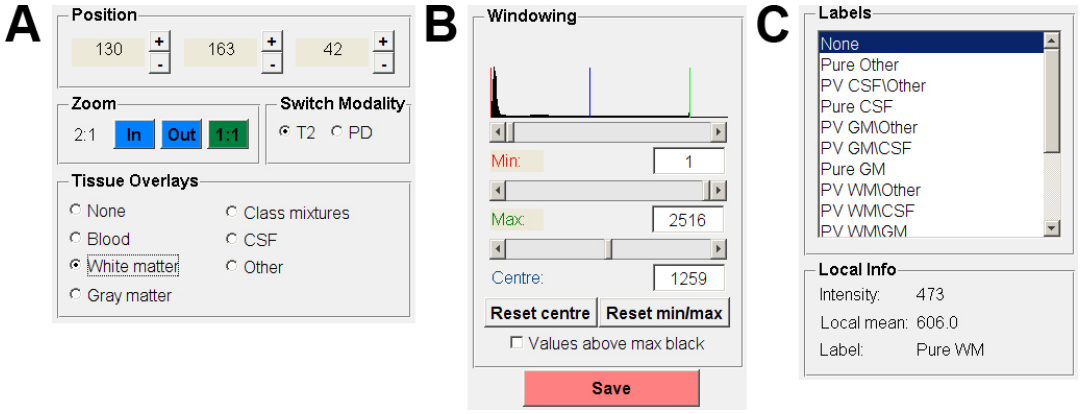

Figure 2.5: **Labelling tool feature panes**. **A**) Navigation and overlays. **B**) Windowing and contrast. **C**) Label selection and information on the highlighted voxel.

currently-selected label, which was chosen from a list since there were relatively few of them. A middle-button click centred the viewport(s) on the chosen voxel and a right-click cleared its assigned label.

Error reporting was accomplished through comprehensive logging as well as informative but plainly-written warning dialogs. Other forms of user feedback included data-dependent control designs – such as sliders for near-continuous values or radio buttons for single selections from multiple categories – as well as a change of mouse pointer graphic for each label brush. Coloured label overlays on the viewports informed the user of the location of specific labels, the content of a specific class, or unlabelled voxels in the current ROI.

2.5 Acquisition of other labelling data

In Chapter 5 the proposed segmentation methods will be tested on the BrainWeb phantom as well as the LBC1936 dataset. This phantom is generated by tissue content maps which are available for download and so formation of the ground truth was trivial. BrainWeb volumes do not contain any blood vessels so the brain parenchyma content mask was created by selecting voxels where the sum of the WM, GM and CSF (and also multiple sclerosis lesions in the case of that phantom) class contents was 1, and then morphometrically closing small holes.

Testing in the following chapters will pit the proposed methods against FAST [Zhang et al., 2001] v4.1, and so the fractional contents of each of the four main tissue classes identified in Section 1.3.1 need to be collected. The following FAST command line options were used once the PD and T_2 volumes had been skull stripped:

- `-S 2`: 2 data channels;
- `-n 4`: 4 tissue classes;
- `--nobias`: No bias field correction;
- `--Hyper 0.55`: Segmentation spatial smoothness (default 0.1);
- `--mixel 0.35`: Spatial smoothness for mixel type (default 0.3);
- No *a priori* probability maps for initialisation;
- Default numbers of iterations were used.

The `Hyper` and `mixel` parameter values were obtained by modifying them in steps of 0.05 from the defaults and examining the change in labelling accuracy in terms of the absolute label error (Eqn 2.14, where lower values are better) and the κ-statistic score (Eqn 2.17, where larger values are better). ROIs 1, 2 and 3 of the LBC1936 dataset were selected for this purpose by virtue of their numerical value and hence can be considered randomly selected. In Fig 2.6 it can be observed that the smallest absolute errors and largest κ-statistic scores over all classes occurred for larger values of `Hyper`. The only result that bucked this trend was ROI 2 where lower values of `mixel` improved the κ-statistic score and higher values of `Hyper` had a slightly detrimental effect; this ROI also posed problems in later chapters due to its coverage of occipital cortex where the WM/GM contrast is low and the cortex is thin, and also since BG

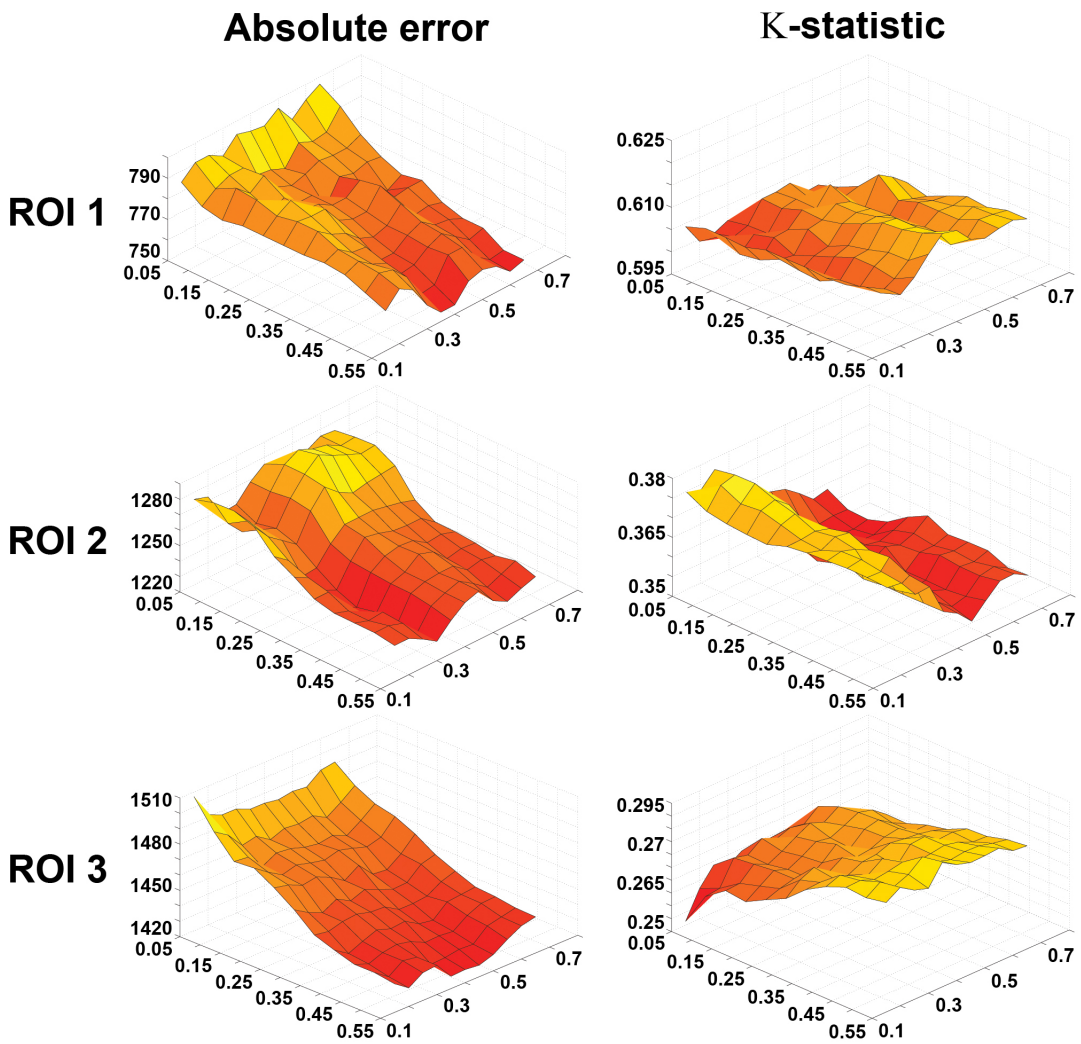

Figure 2.6: **Search for optimum FAST parameters given the surrogate of truth produced by the expert**. Axis names moving anti-clockwise from the vertical axis: total absolute error or κ-statistic score for all classes, `mixel` value, `Hyper` value. Lower values are better for the absolute error, whereas higher κ-statistic scores are desired.

in the falx cerebri did not appear to be labelled properly in the expert-labelled ground truth.

`Hyper` = 0.55 and `mixel` = 0.35 were the best parameters overall, giving small improvements of $[40, 39, 78]$ ($[5\%, 3\%, 5.2\%]$) absolute error and $[0.02, 0.009, 0.023]$ ($[3.4\%, 2.6\%, 9\%]$) κ-statistic score over the worst-perfoming parameter set for each ROI. These parameters also offered the the highest cumulative κ-statistic score and one of the lowest cumulative RMS error scores. Higher values of the parameters were also tested but did not offer a substantial improvement (data not shown), and so in the interests of preventing overfitting these results were not considered.

The unified segmentation method [van Leemput et al., 2003] was also compared against the proposed methods, using a MATLAB implementation provided by Koen van Leemput. The default termination tolerance was used, a 4-class segmentation was specified, and the algorithm was already designed to process multi-channel input. The initialisation process was set to be identical to the proposed methods in this thesis (detailed in Section 5.3.2) and the spatial smoothness of the MRF was estimated and re-calculated during segmentation, and thus did not need to be set.

2.6 Quantitative and qualitative assessment methods

All the evaluation procedures used in Chapters 3-5 will be detailed in this section. Common to all the chapters will be commentary on the results' robustness (stability) – that small variations in the input should result in only small variations in the output – as well as on their reliability – algorithms should behave as expected given a reasonable range of possible clinical input. Non-brain regions identified in each dataset will be excluded from evaluation, in addition to voxels in the LBC1936 dataset that have been assigned non-zero fraction of the *other* class.

The volume-wide measures detailed in this section can be converted to voxel-specific analyses (for example, for the production of plots) by eliminating sums over voxels in the numerator and the appropriate normalising factor of the denominator. The operator # used throughout the following sections denotes a counting operation on a set.

2.6.1 Evaluation of thin structure detection

The labelling of the LBC1936 dataset can be reduced to binary maps B_{true} flagging non-zero content for each class, so that direct comparisons can be made with the binary maps B^{CSF}_{prop} and B^{BG}_{prop} produced by thin structure detection in Chapter 3. The agreement between these maps can be measured using the binary κ-statistic (Eqn 2.3), using the two categories *present* and *none*.

$$P^{agree}_{TSD}(B^{true}, B^{prop}) = \frac{\sum_{v \in \text{voxels}} B^{true}_v = B^{prop}_v}{\#\text{ voxels}} \tag{2.1}$$

$$P^{expected}_{TSD}(B^{true}, B^{prop}) = \frac{\sum_{m \in \{present, none\}} \#(B^{true} = m).\#(B^{prop} = m)}{(\#\text{ voxels})^2} \tag{2.2}$$

$$\kappa_{TSD}(B^{true}, B^{prop}) = \frac{P^{agree}_{TSD}(B^{true}, B^{prop}) - P^{expected}_{TSD}(B^{true}, B^{prop})}{1 - P^{expected}_{TSD}(B^{true}, B^{prop})} \tag{2.3}$$

Listed below are other binary comparison measures can then be compiled, including the true positive (TP) rate (TPR) and false positive (FP) rate (FPR) which are used to create receiver operating characteristic (ROC) curves for parameter tuning and in performance evaluations. The most desirable ROC curve approaches the top left corner of the plot, denoting perfect sensitivity and specificity.

- The **specificity** measured the correctly identified fraction of *none* against all those labelled as *none* in the ground truth. The **false positive rate** is equal to 1 - specificity.

$$\text{Specificity} = \frac{\#\text{TN}}{\#\text{FP} + \#\text{TN}} \tag{2.4}$$

- The **true positive rate**, or **sensitivity**, measured the correctly identified fraction of *present* against all those labelled as *present* in the ground truth.

$$\text{TPR} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FN}} \tag{2.5}$$

- The **positive predictive value** (PPV) measured the correctly identified fraction of *present* against all those labelled as *present* in the proposed method's results.

$$\text{PPV} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FP}} \tag{2.6}$$

- The **negative predictive value** (NPV) measured the correctly identified fraction of *none* against all those labelled as *none* in the proposed method's results.

$$\text{NPV} = \frac{\#\text{TN}}{\#\text{TN} + \#\text{FN}} \tag{2.7}$$

The relationship between a continuous thin structure value T^{prop} and the ground truth content T^{true} (discretised to coarse 0%, 50% and 100% levels of the relevant class from the pure and partial volume labels) was investigated using Pearson's linear correlation coefficient ρ (Eqn 2.8). E is the expected value operator, μ_X is the mean of variable X and σ_X is the standard deviation of variable X.

$$\rho(T^{prop}, T^{true}) = \frac{E\left[(T^{prop} - \mu_{T^{prop}})(T^{true} - \mu_{T^{true}})\right]}{\sigma_{T^{prop}} \sigma_{T^{true}}} \tag{2.8}$$

It was not convenient nor practical for Dr Farrall to describe the local 3D curvature of thin structures in terms of their directions of minimal and maximal curvature, as well as their local shape as plate- or tube-like. Qualitative commentary will therefore include appreciation of the uniformity of their local shape, the congruence of their principal normal directions (Dfn 3.4.1), and the extent and coverage of the binary maps compared to the ground truth tissue content. The latter comparisons will include examining 3D renderings of the thin structures using Simian[14] (as shown in Fig 2.7).

2.6.2 Evaluation of volume resolution enhancement

The fidelity of the volume resolution enhancement process detailed in Chapter 4 has been measured by comparing a volume I^{high}, acquired at a high resolution, with lower resolution volumes I^{vre} which have been resolution-enhanced to match the first. The peak SNR has often been used for the evaluation of super resolution algorithms, though in [Salvado et al., 2006] and [Penney et al., 2004] the mean square error (MSE) was preferred, which is used to derive PSNR. The MSE is a normalised measure which controls against intensity scaling differences, and it penalises large errors between the normalised volumes more harshly than smaller ones. max is the maximum operator.

$$\text{MSE}(I^{vre}, I^{high}) = \frac{\sum_{v \in \text{voxels}} \left(\frac{I_v^{high}}{\max(I^{high})} - \frac{I_v^{vre}}{\max(I^{vre})} \right)^2}{\#\text{ voxels}} \tag{2.9}$$

[14] Available at `http://www.cs.utah.edu/~jmk/simian`.

Salvado *et al.* also used the number of site disagreements ratio (NSDR) to distinguish how many voxels' intensities very poorly matched, using a threshold $0 < \pi < 1$ of the difference based upon a fraction of the dynamic range of the high resolution volume. Again, max is the maximum operator.

$$\mathrm{NSDR}(I^{vre}, I^{high}) = \frac{\sum_{v \in \mathrm{voxels}} \left| \frac{I_v^{high}}{\max\left(I^{high}\right)} - \frac{I_v^{vre}}{\max\left(I^{vre}\right)} \right| > \pi}{\#\ \mathrm{voxels}} \tag{2.10}$$

Slight differences in tissue intensity, noise levels, contrast and bias could exist between the volumes and thus influence the MSE and NSDR; however, edges between tissue regions should remain largely unaffected. The Canny edge detector [Canny, 1986] was employed due to its convenient availability in the MATLAB `edge` function, and since it has been shown that for high SNR its MSE in estimating edge locations is not substantially higher than the lower limit established by the Cramer-Rao bound [Kakarala and Hero, 1992]. The edge map was produced using thresholds tailored to the region under examination, and the standard deviation of the derivative of Gaussian filter set to 2. The mean-square distance figure of merit (FoM, Eqn 2.11), first suggested in [Pratt, 1978] and applied in an edge map comparison framework [Heath et al., 1997], was chosen to score the congruence between the high resolution volume edge map E^{high} and that of the resolution-enhanced volume E^{vre}. The FoM values took the range $[0,1]$ where higher values were more desirable. d is the Euclidean distance to the closest edge in E^{vre} (with a cap of 50) from an edge voxel e identified in E^{high}, and α is a scaling constant for the squared distance which was set to $\frac{1}{9}$. max is defined as the maximum operator once again.

$$FoM(E^{vre}, E^{high}) = \frac{1}{\max(\#E^{vre}, \#E^{high})} \sum_{e \in E^{high}} \frac{1}{1 + \alpha \max\left(50, d(e)\right)^2} \tag{2.11}$$

The entropy of the histogram of the intensities I (Eqn 2.12, examining isosets I_b relevant to each histogram bin) was also investigated by Salvado *et al.* as a measure of the sharpness of its peaks – each one ideally corresponding to a pure tissue intensity. Pure tissue intensity restoration through PVE reduction and smoothing of noise should result in reduced entropy. Entropy was seen as a less reliable measure than the total flow of intensity (Eqn 2.13) between iterations t and $t+1$ for evaluating whether convergence had been reached since the entropy did not monotonically decrease as t

increased. |...| is the absolute value operator.

$$H(I) = -\sum_{b \in bins} p(I_b) \log p(I_b) \tag{2.12}$$

$$TF(I^t, I^{t+1}) = \sum_{v \in \text{voxels}} |I_v^t - I_v^{t+1}| \tag{2.13}$$

Commentary on the resolution enhancement result included examination of the smoothness of intensities in pure tissue regions, non-enhancement of noise, the preservation of thin and small structures and other restorative effects, and the change in crispness of structure boundaries. Any residual bias effects or artifacts were noted and some class content was rendered in 3D using Simian[15] to localise under- or over-estimation in the labelling.

2.6.3 Evaluation of segmentation

Segmentation differences between the ground truth L^{true} and a proposed labelling L^{prop} in Chapter 5 required conversion of the label sets to identical formats amenable to quantification. Pure (assumed 100% content of one tissue class) and two-class partial volume (on average 50% content of each) labels had been produced by Dr Farrall's labelling of the LBC1936 dataset as noted in Section 2.3, whereas the precise class content fractions were known for the BrainWeb phantom (as detailed in Section 2.1). The outputs of the proposed segmentation method and FSL FAST were both continuous values which could be reduced to the same label set as used by Dr Farrall using a partial volume content threshold, when required. If no single class content proportion exceeded this threshold then the two classes with the greatest content were selected as the partial volume constituents; otherwise a pure label was selected.

The absolute label error (Eqn 2.14) and κ-statistic score (Eqn 2.17) have been employed for the quantitative assessment of labelling error. In [Bricq et al., 2008] the root mean square (RMS) error between the class proportion label vectors for continuous quantification was used instead, but this penalises larger errors more heavily than small ones and thus is not directly indicative of the volume of voxels in error. The absolute error is sensitive to small fluctuations between two labellings of the same volume, particularly if one has been discretised, and so κ measures only the correctness of the label matching in comparison to chance matches but not the magnitude of the

[15] Available at `http://www.cs.utah.edu/~jmk/simian`.

errors. The maximum value of κ is 1, signifying perfect agreement; 0 indicates a similar accuracy level as if the labels in L^{prop} were randomly chosen; and the minimum is -1, which means L^{prop} is completely inaccurate. The interpretation of Landis and Koch [Landis and Koch, 1977] stated that a result over 0.8 should be considered almost perfect, but κ will be higher when there are fewer categories [Sim and Wright, 2005], so this factor must be appreciated when the cardinality of the label set M changes. Calculation of κ can also be performed for a single class k with a much smaller matching matrix. $|...|$ is the absolute value operator.

$$ABS(L^{prop}, L^{true}) = \sum_{v \in \text{voxels}} |L_v^{true} - L_v^{prop}| \tag{2.14}$$

$$P^{agree}(L^{prop}, L^{true}) = \frac{\sum_{v \in \text{voxels}} L_v^{true} = L_v^{prop}}{\#\text{ voxels}} \tag{2.15}$$

$$P^{expected}(L^{prop}, L^{true}) = \frac{\sum_{m \in M} \#(L^{true} = m).\#(L^{prop} = m)}{(\#\text{ voxels})^2} \tag{2.16}$$

$$\kappa(L^{prop}, L^{true}) = \frac{P^{agree}(L^{prop}, L^{true}) - P^{expected}(L^{prop}, L^{true})}{1 - P^{expected}(L^{prop}, L^{true})} \tag{2.17}$$

Qualitative evaluation of the results will present comments on aspects which are subjective or difficult to quantitatively measure including the uniformity of labelling in the presence of noise, the correct classification of thin or small structures and pathology, and the displacement or smoothing of boundaries. Any bias effects, pathology or areas of low contrast in the volume that may negatively impact the results will be noted.

2.6.4 Measurement of algorithmic efficiency

All computation times are listed for a AMD Athlon 64 X2 Dual Core 4800+ PC with 4Gb of RAM, running the Windows 7 64-bit operating system. MATLAB was selected as the main software development platform due to its ease of prototyping and range of visualisation tools, though as an interpreted language it can execute programs slowly. In future chapters it has been noted where external code sources have been used or a few blocks of computationally-intensive code have been rewritten in MATLAB-compatible C `.mex` files and compiled in order to speed up processing.

The processing times of algorithms have been captured over 3 repetitions (using the same parameters and data). Every non-essential program or service was closed and the machine was isolated from network connections in order to remove variability from these sources. Relevant parameter values, the dataset size and the number of iterations taken to reach convergence have been noted alongside the running time. While processing time is dependent on implementation efficiency and hardware specifications, it can also be bounded by a function of its parameters and the input data size. Estimates of the order of computational complexity will also be presented in the following chapters.

2.7 Discussion

The testing corpus incorporates a range of dual-channel brain MRI data to enable proper assessment of the accuracy of the proposed methods: it has been acquired from a young volunteer with healthy tissue under ideal scanning conditions, an aged cohort which should challenge segmentation algorithms with low contrast and degenerated tissues, and a software phantom allowing precise quantification of content estimation errors as well as testing robustness to multiple sclerosis pathology. More thorough testing and parameter training could be accomplished with a larger set of labelled data including volumes acquired from different scanners with a larger variety of subjects (in terms of race, age, gender, brain size, amount of degenerative or pathological tissue and the amount of motion they make inside the scanner). Plans for multiple-expert labelling and repeated labelling averages have been outlined in the Conclusions (Chapter 6) since unfortunately there was only one expert available and no opportunity for Dr Farrall to repeat his labelling of the LBC1936 dataset. The granularity of the label set available was poor but given the poor contrast of some regions of the LBC1936 dataset then multiple possible fractional content levels of two-class mixtures would increase the chance for misclassification, as well as increase the time needed to select the appropriate label for a voxel.

During the selection of pre-processing tools, it was surprising that the bias correction tools tested did not consistently produce a beneficial effect. Positive reviews in two round-ups [Zaini et al., 1999][Arnold et al., 2001] were cited to support the use of BFC and N3, but the accuracy of such algorithms on real brain volumes is difficult to quantify: there is a reliance upon subjective appreciation of the intensity histogram and changes in intensity uniformity since no ground truth correction can be acquired.

Speculating upon the reasons for their failure, poor performance of the single-channel soft segmentation model used by BFC may have been caused by low contrast between some tissue classes and structures, and similar issues manifesting as wide peaks on the intensity histogram may have hindered N3. A new bias field correction approach is presented in Section 6.5.3 which is similar in philosophy to BFC, but uses both T_2- and PD-weighted data and can be integrated with the work in this thesis.

The software tool for assisting manual labelling of the LBC1936 dataset ROIs was produced since there is no viable *in vivo* alternative to manual painting of labels by human experts for producing the ground truth. An important design decision was to only allow the labelling of individual voxels and not to assist the user though region selection techniques such as active contour evolution or graph theory-based livewire [Falcao et al., 1998] algorithms since this would not encourage full investigation of each voxel and may even promote carelessness. Evaluation of the tool was undertaken by interviews with Dr Farrall; ideally other radiologists that were not involved in the design would also be consulted, and quantitative testing of improvements in labelling speed and accuracy would be performed. Some feedback could also be obtained by placing the tool in the public domain, and it has been made part of the RESCUE segmentation package (with details listed in Section 6.6). Some mundane features which could improve its utility include reporting of the physical volume of labelled classes and the estimation of class intensity distribution parameters to assist intensity windowing (a method for the estimation of Gaussian-distributed class parameters is detailed in Section 5.3.5).

3D visualisation of the data within the tool is an intriguing issue since it could improve appreciation of structure boundaries and overall shape compared to 2D viewing restricted to the three orthogonal planes. However, rotation and manipulation of 3D objects can be complicated, and viewing will be confounded by transparency issues once the structures of interest are deep inside the brain. Furthermore, labelling will still be restricted to the 2D views due to the difficulty of selecting voxels in 3D. Rendering using VTK[16] or Simian[17] may be sufficient; the latter tool has been used throughout this thesis to visualise 3D structures due to its excellent manipulation tools and ability to highlight structures below the surface of objects. The colour, hue and transparency of rendering in Simian are determined by a transfer function working with the zero- to second-order derivatives of the data, and the operation of the function can be altered

[16] Available at `http://www.vtk.org/`

[17] Available at `http://www.cs.utah.edu/~jmk/simian`.

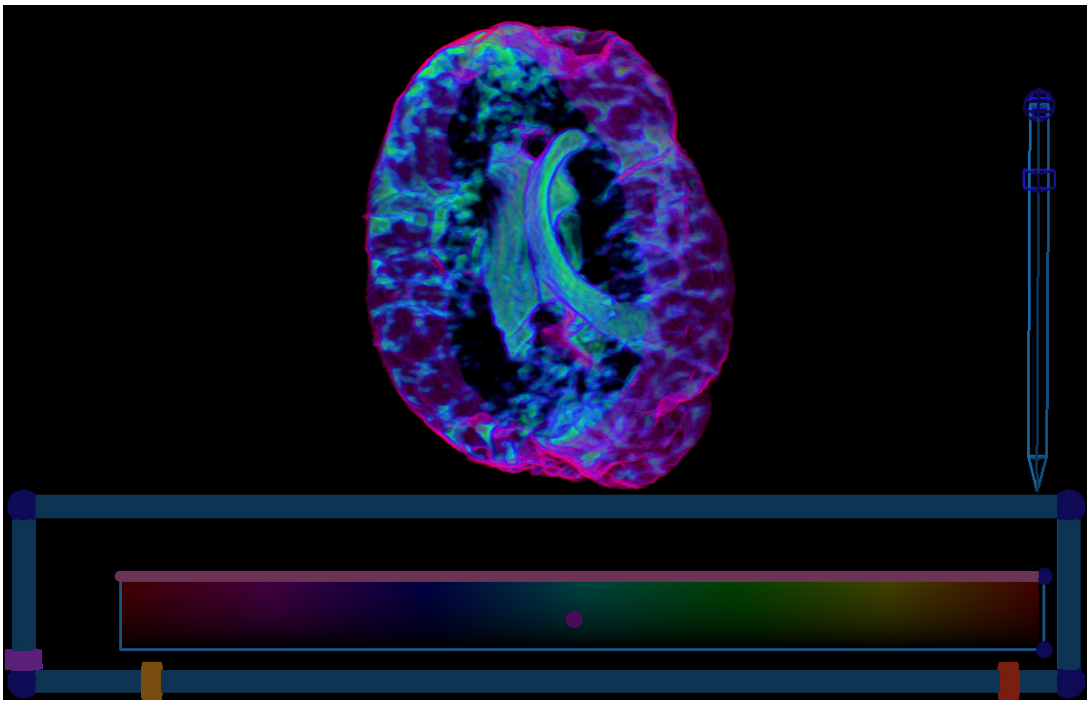

Figure 2.7: **3D visualisation using Simian**. T_2 data from a volume in the LBC1936 dataset is shown, with a transfer function widget (using a rainbow look-up table) defined to highlight CSF. The curved lateral ventricle is contrasted in green-blue below the surface of the parenchyma.

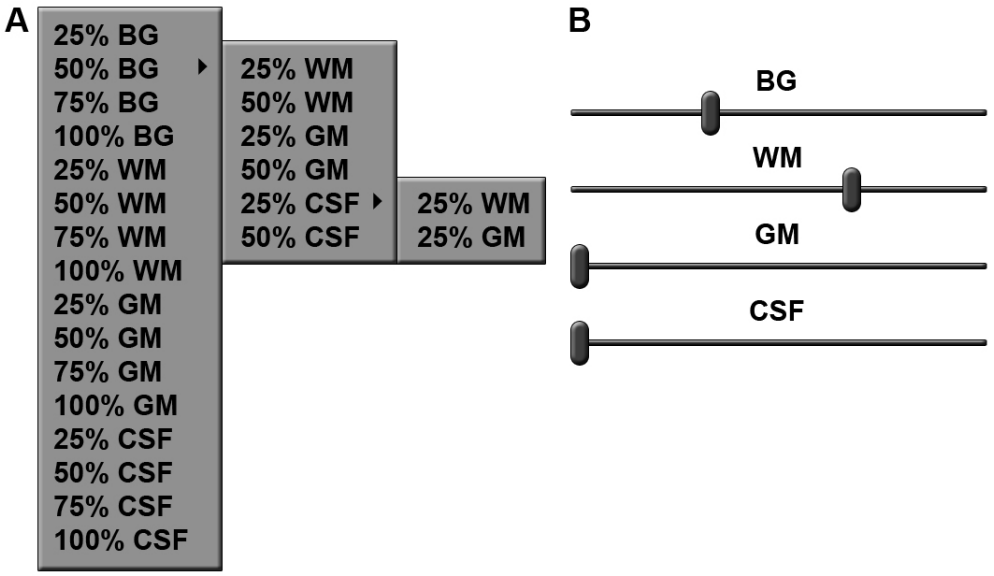

Figure 2.8: **Proposed changes to the label selection interface with increasing cardinality of the label set**. **A**) A hierarchy of context menus, with a number of levels equal to the maximum number of classes that can contribute to a mixture, could save space and allow for fast label selection. **B**) A continuous selection interface with a slider for each class may be more appropriate when the fine-grainedness substantially increases.

simply using an interactive widget.

Two interface elements of the tool are obvious candidates for further development. Firstly, the labelling tool lacks a ROI definition mode for the $50 \times 50 \times 1$ voxel square axial-plane ROIs employed in this thesis. Structures of interest may be more fully and efficiently enclosed (reducing the labelling workload) using free-form boundaries or regular 3D shapes defined using a small number of parameters – excellent ROI definition tools have been developed for MIPAV and ITK-SNAP which could be replicated. Secondly, the label selection interface demands a more intuitive and compact selection system once the list of labels grows, so that the user is not required to search and scroll through a list where many labels could be hidden from the current viewport. The ordering of the labels and the method of selecting a label may be more intuitive and efficient with a tiered context menu system (Fig 2.8A), or a set of sliders for very fine-grained label sets (Fig 2.8B).

The evaluation methods incorporate well-established methods of quantifying differences between pairs of volumes' intensities, binary maps and multi-category maps that are necessary for determining the accuracy of the proposed methods in each chap-

ter. The edge detection function provided with MATLAB operated in 2D, and so a simple improvement to the quality of the edge maps could be made by using 3D processing which is more appropriate for volumetric data. Entropy of the intensity histogram has been used to judge the amount of pure tissue intensity restoration, though both bias field effects as well as local fluctuations in pure tissue intensities as residue from smoothing noise may combine to reduce this measure's usefulness. Further use could be made of the matching matrix used in the calculation of the κ-statistic to examine the precise examine the label confusion in sub-volumes; this type of analysis has only been performed from visual examination of the estimated class content plots.

Qualitative assessment of accuracy and robustness is also necessary since abnormalities (such as low contrast, bias field effects and pathology) and some global trends can be difficult to quantify, and peculiarities or mistakes in the ground truth must be noted. This last step is crucial since the ground truth is in fact a *surrogate of truth* (using the terminology in [Udupa et al., 2006]): labelling decisions are subjective and affected by factors including local light levels, expert knowledge of the anatomy, skill at using the labelling tool, and other factors affecting current rater performance. Some 3D visualisation has been performed, but the principal normal directions (Dfn 3.4.1) of thin structures could be visualised in their entirety using a tool designed for DTI visualisation such as 3D Slicer[18]. All the qualitative commentary could be improved by obtaining commentary from another radiologist.

The 64-bit computer on which processing was performed did not pose addressable memory constraints, but more powerful multi-processor hardware will allow processing to complete in much shorter timescales when measuring efficiency. A more radical change would be to use a computing cluster or grid such as EDDIE[19]. In future chapters parallelisable components of the algorithms have been identified and more computing power will allow the processing of larger volumes, more complex optimisation as well as restarts, and parameters can be set to values which will likely increase accuracy at the penalty of lengthening the processing time.

[18] Available at `http://www.slicer.org/`.

[19] Present at the Edinburgh Compute and Data Facility at `http://www.ecdf.ed.ac.uk/`, and available to Informatics researchers at the University of Edinburgh.

2.8 Summary

In this chapter the procedures for data and labelling acquisition (Sections 2.1, 2.3 and 2.5), pre-processing (Section 2.2) and evaluation (Section 2.6) have been detailed for use in future chapters. The evaluation strategy for fulfilling the key metrics listed in Section 1.3.5 incorporated a host of established quantitative measures, and qualitative guidelines were stated for where quantitative analysis was not feasible. In addition to examining the accuracy of the proposed methods as well as the robustness through the range of data acquired, other performance indicators gathered included statements on algorithmic efficiency and measurement of the processing time. To enable comparisons with other segmentation tools then in Section 2.5 the parameters set for processing data using FAST [Zhang et al., 2001] and the unified segmentation method [van Leemput et al., 2003] were detailed.

In order to assist manual labelling of the LBC1936 dataset and assist analysis of accuracy, in Section 2.4 a software tool was designed and implemented in MATLAB since no currently-available tool was found completely suitable for soft labelling of volumes. Unfortunately only one expert was available to produce a single labelling of the LBC1936 dataset so more labelled data, more repetitions and more expert raters would have been desirable.

Chapter 3

Thin contrasting structure detection and anisotropy-preserving filtering

Image features – such as anatomical structures in a brain MRI volume – can be visualised at different spatial scales (Fig 3.1). In Section 1.3.5 it was apparent that brain tissue segmentation tools such as FAST [Zhang et al., 2001] are generally accurate when processing well-contrasted uniform tissue regions at multiple-voxel scales, and much of their success can be attributed to neighbourhood label interaction schemes that promote local homogeneity in the presence of noisy intensities: further details on spatial regularisation using MRFs can be found in Section 5.1. The spatial filters used to weight the importance of neighbouring voxel data – ignoring label interaction scores – are usually isotropic in form (defined below), though the accuracy for anisotropic structure processing is of the most interest in this chapter.

Definition 3.0.1 A filter F of dimension N is **isotropic** (or radially symmetric) if $\forall i, j \in \mathbb{R}^N, (\text{distance}(c,i) = \text{distance}(c,j)) \implies (|F_i - F_j| \leq \varepsilon)$. c is the position index of the central element of the filter (usually the zero vector), i and j are other position indices, $\implies$ is the logical implication operator and ε is a threshold close to or equal to zero. In the case of Gaussian filters, to satisfy this property then their covariance matrix Σ should be close to a form where the diagonal elements are equal (assuming unit lengths in each spatial dimension are equal) and the off-diagonal elements are zero.

A structure S consisting of binary structure elements and possessing 3 spatial dimensions is roughly isotropic (or almost spherical) if $\forall i, j \in \mathbb{R}^3, [\text{border}(i) \wedge$

border(j)] $\Longrightarrow$ ($|$distance(c,i) − distance$(c,j)| \leq \varepsilon$). Again, c is the position index of the central element of the structure, i and j are other position indices, and ε is close to or equal to zero.

If the isotropic condition does not hold, then the filter or structure is **anisotropic**.

Anisotropic structures possess one or two dimensions which are thinner than the other(s). The main effects of performing isotropic filtering on anisotropic structures – which are small relative to the spread parameters of the filter – include blurring and diminished contrast (investigated in more detail in Section 3.2). At typical clinical MRI resolutions two anisotropic structure types (described in Section 1.2.3) are particularly badly affected by the isotropic filtering which is relevant for many WM and GM structures:

- **Cortical sulci**, which are CSF-filled fissures between folds in the cortical GM, and appear very bright in T_2-weighted data;
- Connected networks of tubular **blood vessels** that are well-contrasted in PD-weighted data.

The methods described in this chapter for detecting these structures' presence, shape, scale and orientation enable processing which is dependent on these features: information flow can be promoted along the longer dimensions of thin anisotropic structures in order to preserve their high spatial frequencies. This adaptive filtering has been applied to enhance the resolution and smoothness of the MRI data (Chapter 4) and to segment brain tissue (Chapter 5) while preserving these thin structures. The philosophy of this approach is similar to adaptive anisotropic diffusion smoothing [Knutsson et al., 1983][Perona and Malik, 1990] which seeks to smooth along, rather than across feature boundaries. Detecting these structure features involves a novel smoothing operation to ensure fuller coverage of the maps.

This chapter begins with an examination of the behaviour of single-scale isotropic filtering in the presence of these particular thin structures, a brief review of methods which have previously been employed for their detection and adaptive processing, and a description of the multi-scale vessel enhancement filtering approach presented in [Frangi et al., 1998]. Differential geometry and scale spaces have been introduced

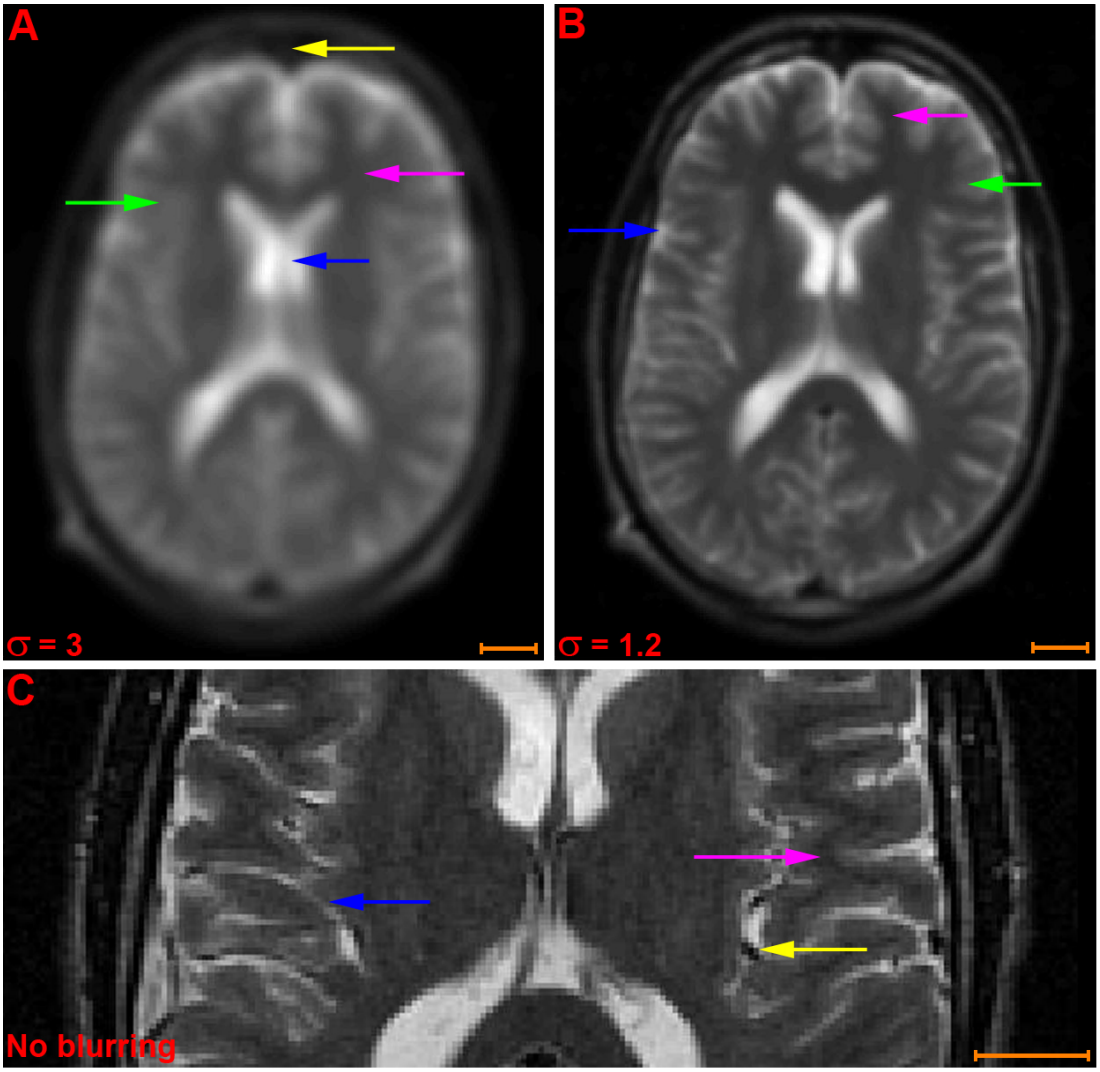

Figure 3.1: **Features of structural MRI of the brain visible at different spatial scales**. Axial T_2 data from the young brain dataset are shown. Pink arrows denote WM structures; blue arrows denote CSF structures; green arrows denote GM structures; and yellow arrows denote blood and background structures. Scale bar (orange) width: 10 voxels. **A**) Gaussian blurring with $\sigma = 3$. At the largest scales, large white matter tracts, the skull, the ventricles and larger GM regions can be seen. **B**) Gaussian blurring with $\sigma = 1.2$. At a smaller scale the cortical sheet, many smaller white matter regions and larger CSF-filled sulci are observable. **C**) No blurring. At the smallest scale blood vessels, thin white matter regions between gyri and thinner sulci are visible.

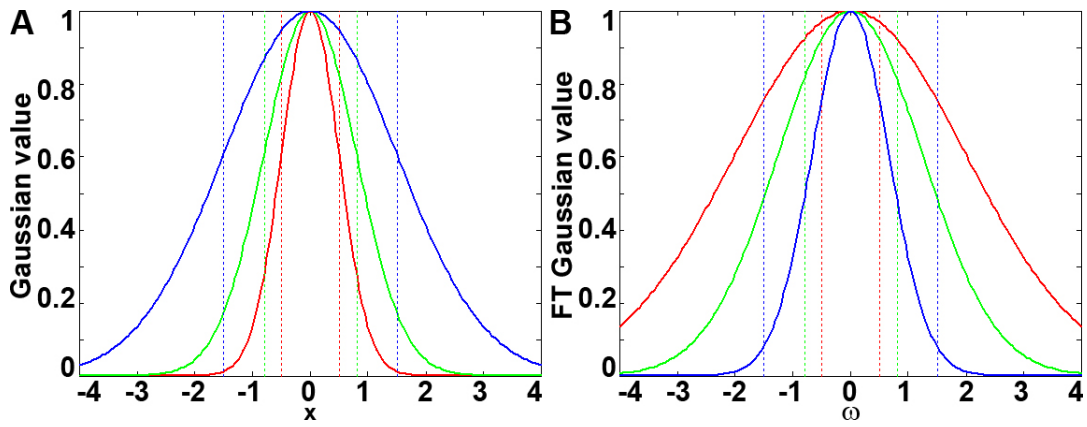

Figure 3.2: **Fourier transforms of 1D Gaussian functions**. The Gaussian function used is $e^{-\frac{x^2}{2\sigma^2}}$ and its Fourier transform is $e^{-\frac{\omega^2\sigma^2}{2}}$. The standard deviation of each function is marked with a dashed line of the appropriate colour. **A**) Gaussian functions. **B**) Fourier transforms of Gaussian functions plotted in **A**. **FT**: Fourier transformed.

beforehand as necessary pre-requisite knowledge. The modifications necessary for application of the vessel enhancement filtering approach to T_2-PD dual-channel brain volumes have then been formally described with some discrete implementation details. Next, demonstrations using the young brain data and the expert-labelled LBC1936 dataset have been performed, and the results have been discussed (mainly qualitatively) in the context of the literature. Some improvements have been suggested for more accurate, complete and efficient processing as well as for error correction, which are detailed either in this final section or in the Conclusions (Chapter 6) of this thesis.

3.1 Introduction to Gaussian smoothing and scale spaces

Smoothing involves the blurring of high spatial frequency features in N-dimensional data V; volumetric data is considered in this thesis, and so $N = 3$. Commonly, the operation involves convolution[1] of V with a low-pass filter such as a Gaussian function g (Eqn 3.1). Σ is the covariance matrix describing the spread of g – the decrease of the output value with increasing distance from the central position (1D examples shown in Fig 3.2A) – in each dimension. Since the Gaussian function decays quickly it is reasonable to define a finite filter window based upon Σ, outside of which the filter values are assumed to be zero. The size of the window denotes the filter's applicability range, defined in Eqn 3.1 as a, and its magnitude determines the number of calculations

[1]Denoted by the symbol $\otimes$.

performed during convolution in a discrete implementation.

$$V_{smooth}(V,\Sigma) = V \otimes g(\Sigma) = \int_{y \in V} \int_{x \in a} V(y-x) \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-(x\Sigma^{-1}x^T)} dxdy \qquad (3.1)$$

The frequency response of a Gaussian function can be appreciated by taking its Fourier transform, which is also Gaussian (Fig 3.2B). Taking a low response threshold τ for its assumed 1D functional form $e^{-\frac{\omega^2\sigma^2}{2}}$, then structures which have a spatial frequency greater than $\sqrt{-\frac{2\ln(\tau)}{\sigma^2}}$ can be considered blurred after smoothing. A more general statement can be made in that there is an inverse relationship between the filter spread and its degree of preservation of high spatial frequency structures.

Isotropic Gaussian filtering with a range of different Σ creates a scale space representation of a dataset [Witkin, 1983]; when Σ is the zero matrix then the output of Eqn 3.1 is original volume, which is the highest effective resolution where the finest structures can be observed but also where the power of noise is highest. As the diagonal elements of Σ increase then only lower-frequency (thicker, wider or longer) structures retain their appearance since high frequency structures, including noise, are blurred by the low-pass design of g. The usefulness of isotropic Gaussian filtering for such multi-scale analyses lies in the functions' fulfilment of scale space axioms[2] including non-enhancement of local extrema[3], scale invariance and rotational invariance. Perhaps the most important property of scale spaces for clinical image processing is that smoothing at high Σ does not create new structures that do not correspond to simplifications of corresponding structures which are more clearly visible at low Σ.

3.2 Behaviour of smoothing in the presence of thin structures

The contribution of intensities from inside a thin (anisotropic) contrasting structure can be poor when isotropic or wrongly-shaped or -oriented anisotropic Gaussian smoothing is performed. In these cases the applicability range of the smoothing filter extends beyond the structure's boundaries in its thin dimensions (depicted in Fig 3.3A), and in a discrete implementation this effect can be magnified at low resolutions since border voxels make up a significant proportion of these structures' volumes. These voxels are

[2]For a more in-depth description of scale spaces the reader is referred to work on this topic by Lindeberg [Lindeberg, 1993].

[3]Also known as the minimum-maximum principle.

likely to be affected by PVE and so the filter coverage may include substantial contributions from other structures. A real example is shown in Fig 3.3B-E: the preservation of the contrast of a blood vessel (highlighted in red) after smoothing is affected by the degree of similarity between the local structure orientation and the orientation of the longest axis of the 3D Gaussian smoothing filter used. This result advocates adaptive smoothing with anisotropic filters which are closely shaped to local structure, an approach which has been employed in several different application domains (i.e. [Haglund, 1992][Svensson, 2008]) and is an active topic of research.

Assuming that sections of thin anatomical structures can be modelled approximately and simplistically as Gaussian functions, anisotropy confers the sections with the ability to realise tube and plate primitive shapes, whereas more isotropic structures can be considered blobs (all defined below and pictured in Fig 3.4).

Definition 3.2.1 (Primitive shape covariances) A **plate**-shaped 3D structure has high spatial frequency – or equivalently low thickness – in one direction, forming the normal to two directions of low spatial frequency. Plates possess a covariance matrix which is similar to the following form, where $large > small > 0$.

$$\Sigma_{plate} = \begin{pmatrix} small & 0 & 0 \\ 0 & large & 0 \\ 0 & 0 & large \end{pmatrix} \tag{3.2}$$

A **tube**-shaped 3D structure has low spatial frequency – or equivalently high thickness – in one direction, forming the normal to two directions of high spatial frequency. Tubes possess a covariance matrix which is similar to the following form, where $large > small > 0$.

$$\Sigma_{tube} = \begin{pmatrix} large & 0 & 0 \\ 0 & small & 0 \\ 0 & 0 & small \end{pmatrix} \tag{3.3}$$

A **blob**-shaped 3D structure has similar spatial frequency in all directions, and so possesses a covariance matrix similar to the identity matrix multiplied by a scalar value.

Acquisition of the information listed below can provide a good estimate of – in

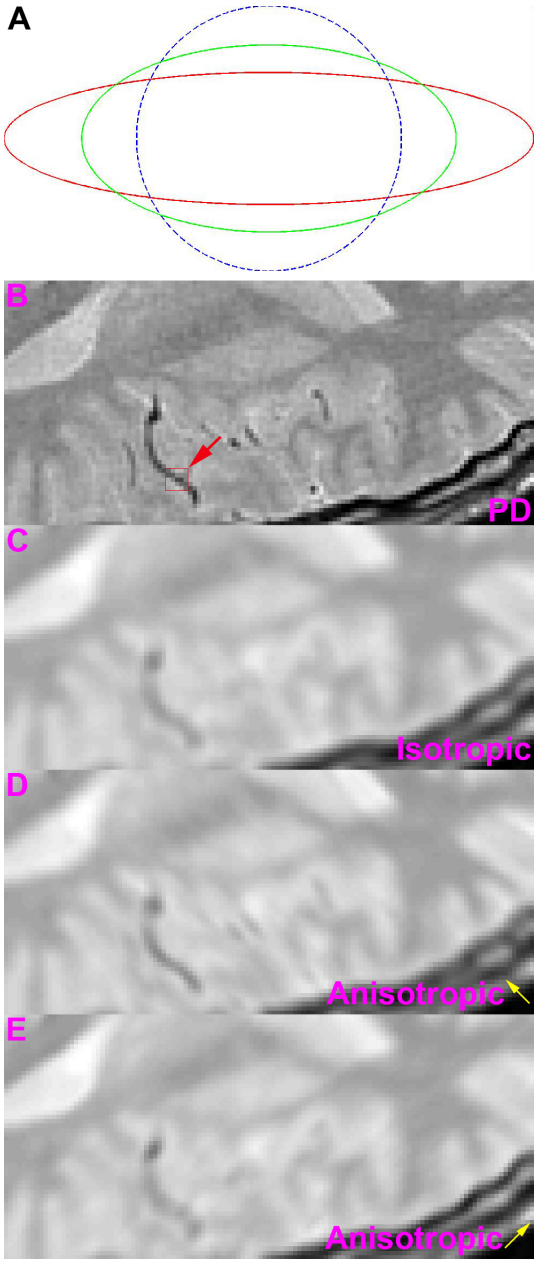

Figure 3.3: **Effects of isotropic and oriented anisotropic smoothing on anisotropic structures**. **A**) Plotted are the area-normalised covariance matrices of a thin structure modelled with an anisotropic Gaussian (in red), an isotropic Gaussian filter (dashed in blue), and their convolution (in green). **B**) Axial PD data from the young brain dataset. The red arrow points to an anisotropic structure – a blood vessel proceeding in the axial plane. **C**) 3D isotropic Gaussian smoothing of **B**. **D**) Non-adaptive anisotropic smoothing of **B**, with the filter covariance oriented along the vessel. **E**) Non-adaptive anisotropic smoothing of **B**, with the filter covariance oriented perpendicularly. Yellow arrows denote the orientation of the longest principal axis of the 3D Gaussian filter covariance for anisotropic smoothing.

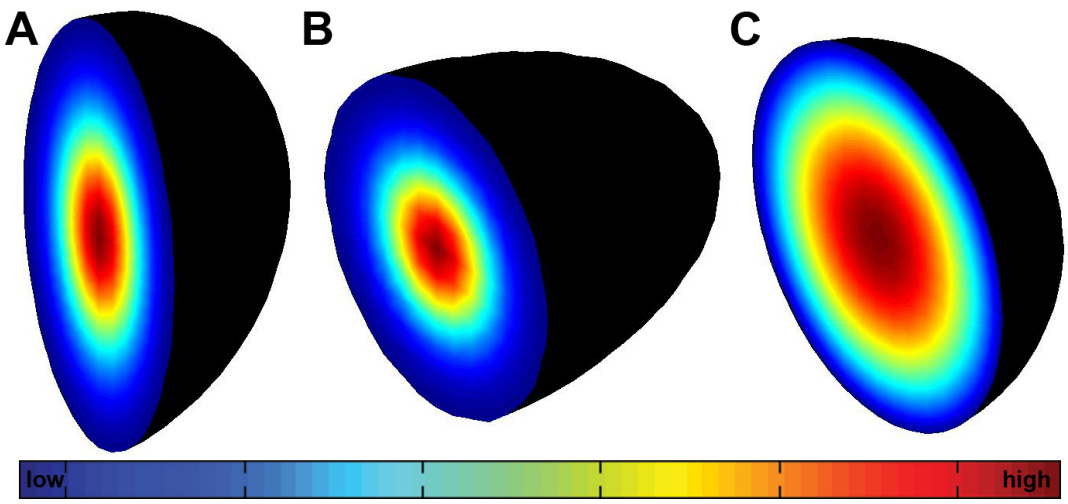

Figure 3.4: **3D Gaussian functions used as models for sections of isotropic and anisotropic structures**. Cross-sections through the centres are shown, and the ratio between the diagonal elements of their covariance matrices (defined in Dfn 3.2.1) is $large : small = 2 : 1$. **A**) $g(\Sigma_{plate})$. **B**) $g(\Sigma_{tube})$. **C**) $g(\Sigma_{blob})$.

other words, ideally positively correlate with – the local structure covariance matrix. In this way, similarly-designed anisotropic filtering applied adaptively to thin structures can help to preserve their structural anisotropy.

- The **shape**, as a plate or tube (Dfn 3.2.1);
- The **ratio between variances** $small$ and $large$;
- The **orientation**, later applied as a rotation transformation to the filter.

3.3 Previous approaches for thin contrasting structure detection

Characterisation of two particular thin structures – blood vessels and cortical sulci – in terms of their length, thickness and other shape parameters has proved important for understanding normal anatomical variability as well as studying developmental disorders, pathology and the effects of ageing[4]. In this section differential geometry-based techniques for detecting contrasting anisotropic structures will chiefly be examined since they form the foundation for the methods proposed in Section 3.4. In [Lopez

[4]The reader may wish to examine references [Cocosco et al., 1997] and [Kirbas and Quek, 2003] for more comprehensive literature reviews on blood vessel and cortical sulci detection.

et al., 1999] several desirable properties for thin structure enhancement filtering were suggested, which are also relevant to the task presented in this chapter:

- There should be good contrast with surrounding areas;
- Similar values should be maintained in the direction along a particular structure;
- The greatest values should found at the core of anisotropic objects[5];
- They should be robust to small image perturbations, such as noise.

The first candidate method for identifying vessels and sulci is phase congruency [Kovesi, 1999] which seeks high phase ordering in the Fourier transform of data. In a very simple example representative of step edges, a Fourier-transformed square wave will have all its sine waves in phase at the points where steps occur in the signal. Elements of phase congruency have been deemed crucial to visual perception [Openheim and Lim, 1981] and it can be generalised to also discriminate features such as lines and corners. Importantly, in the context of brain MRI the measure may be superior to gradient-based edge detection methods (i.e. [Marr and Hildreth, 1980][Canny, 1986]) since it is robust to changes in intensity contrast: the calculation of phase congruency is independent of the signal magnitude, and a basic 1D implementation is shown in Eqn 3.4. A_n is the amplitude of the n^{th} Fourier component of the signal, $\theta_n(x)$ is the local phase of this component at position x, and $\overline{\theta}(x)$ is the amplitude-weighted, mean local phase angle of all the Fourier components at x which maximises this equation. Effectively, phase congruency finds a minimum of the weighted variance of local phase angles relative to the average local phase.

$$\mathrm{PC}(x) = \max_{\theta(x)\in[0,2\pi]} \frac{\sum_n A_n \cos\theta_n(x) - \overline{\theta}(x)}{\sum_n A_n} \tag{3.4}$$

In [Linguraru et al., 2003] phase congruency was applied to a structural T_2 volume with the objective of detecting the thin contrasting features corresponding to cortical sulci, and in Fig 3.5 a 2D implementation[6] of phase congruency was applied to a T_2-weighted slice from the LBC1936 dataset using four spatial scales and six orientations. It was observed that many of the features identified did correspond with sulci, but their extent was generally overestimated, and there were many other false positive results (marked with red arrows). At these small scales patches of noise were problematic and

[5]This line or plane is usually called the medial axis.

[6]Code provided by Peter Kovesi at `http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/`.

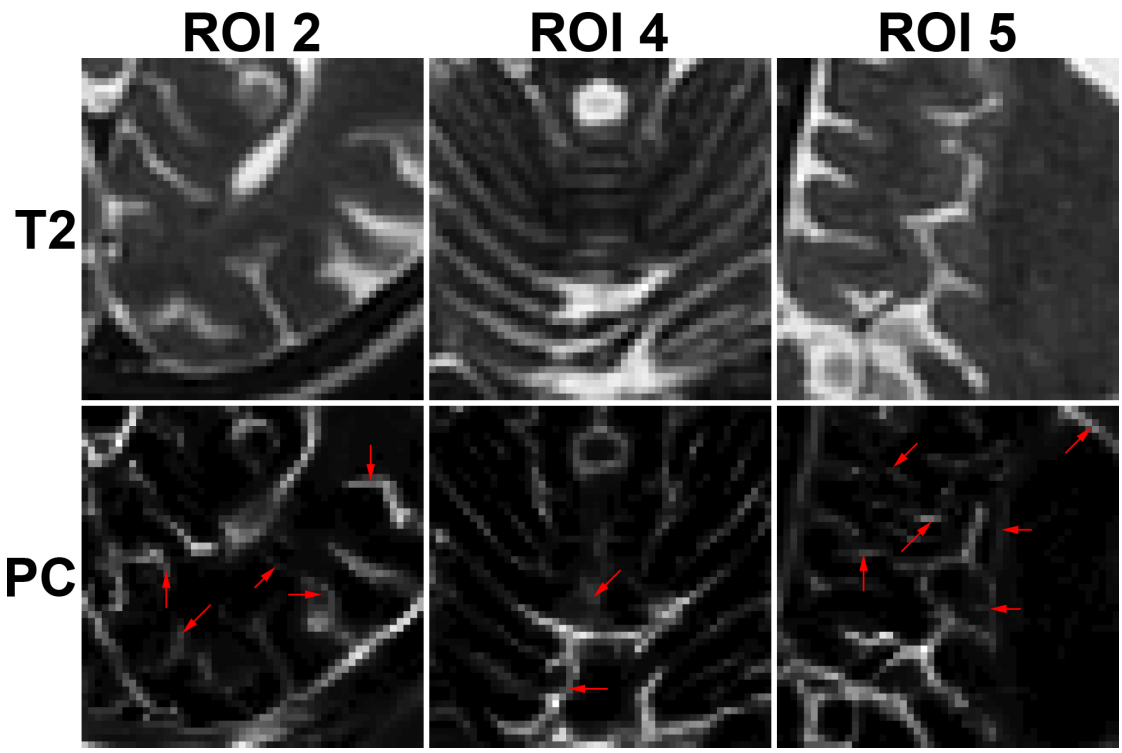

Figure 3.5: **Detection of cortical sulci using phase congruency**. **A**-**C**) Regions of an axial slice of a T_2 volume taken from the LBC1936 dataset. **D**-**F**) Phase congruency values (brightness denotes higher values) obtained from **A**-**C**. Red arrows mark features identified by phase congruency that do not correspond to cortical sulci. **PC**: phase congruency values.

in some cases they managed to cause branching of the measure between thin structures; furthermore, running the code with a variety of parameter values revealed that the specificity was extremely sensitive to the noise estimation procedure. Consequently this method has not been chosen for the proposed methods in Section 3.4, but it has been examined again in the Discussion (Section 3.6).

Thin structure detection has often involved segmentation processes in the literature, with techniques ranging from simple thresholding and morphology [Royackkers et al., 1999] to more advanced methods based upon moving windows [Yi and Ra, 2003] that promote continuity along their lengths and provide resistance to noise, bias and PVE. However, such simple appreciation of intensities does not consider any notion of local shape – which is necessary to discriminate thin structures from others with the same constituent tissue type – and so additional structural modelling must be employed. Other discriminative models for detecting particular cortical sulci have typically employed a wide range of basic features including intensity, gradients, location

in stereotaxic co-ordinates and Haar filter results (i.e. [Zheng et al., 2006]) but suffer from the same failing. Training multi-feature systems may also require restrictive amounts of positive- and negative-labelled data, and the adaptiveness of the method is important: over-specified models may not transfer well between volumes which have been acquired at different times or from different patients or scanners.

An explicit cortical model can also be used to model gyri and extract sulci. Freesurfer [Dale et al., 1999][Fischl et al., 1999], a set of automated tools for reconstruction of the brain's cortical surface from structural MRI data, incorporates classification of parenchymal voxels into WM and non-WM classes using intensity as well as neighbourhood label context (from both immediate neighbours and along the local plane of least intensity variance) and prior probabilities from a labelled atlas. Subsequently a connected-components procedure establishes the WM content, which is then surrounded by a surface and the intensity gradient and curvature are examined normal to this surface to produce a GM/CSF boundary further away. As FreeSurfer uses WM as the base deformable surface then topological errors at the GM/CSF boundary due to mistakes in the segmentation (such as noise labelled as GM "bridging" a thin sulcus) can easily be corrected manually; however, the quality of the WM/GM contrast – much lower than with CSF in dual-echo T_2-PD volumes – is a key contributor to determining the accuracy of sulcal extraction.

The final group of candidate methods investigated generates a description of local shape naturally through higher-order intensity derivatives. Differential geometry can investigate the local behaviour of intensity changes in a volume V represented at scale-space level s through the Taylor expansion in the neighbourhood of a point x (truncated at the second-order term in Eqn 3.5). The first-order differential quantity is the gradient $\bigtriangledown$, providing a vector pointing to the path of steepest intensity ascent, and the second derivatives are contained in the symmetric Hessian matrix H. $\bigtriangledown$ and H are universal derivative calculators: they allow the efficient computation of derivatives in any direction through multiplication with that direction's vector, and they do not require discretisation of the orientation space.

$$V(x+\Delta x,s) \approx V(x,s) + \delta x^T \bigtriangledown V(x,s) + \delta x^T H[V(x,s)]\delta x + \ldots \tag{3.5}$$

The appropriate first or second derivative of an isotropic Gaussian filter g with variance s in each dimension is used to create $\bigtriangledown$ and H, which then allow investigation of the local structure over different scale-space levels. Intensity derivatives are decreasing functions of scale, and so scale-normalised and γ-parameterised Gaussian

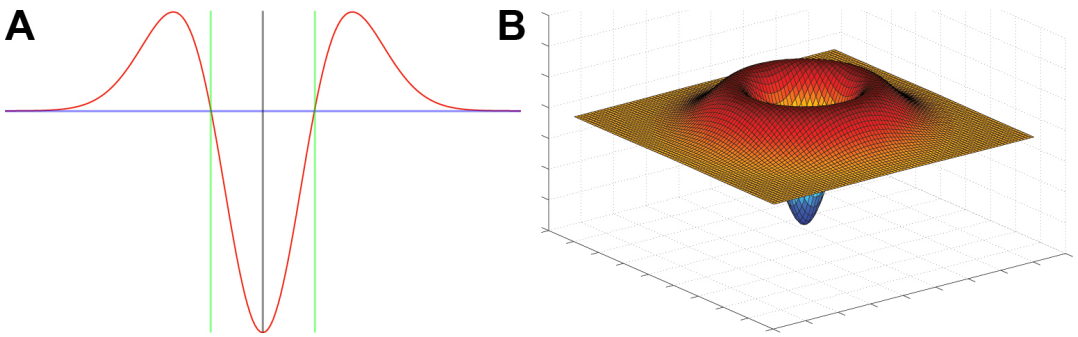

Figure 3.6: **Second derivative Gaussian probe kernels**. Height denotes function values. **A**) 1D function. The faint blue line denotes zero value of the function, the gray line shows the central element (zero) position, and the green dotted line indicates the σ position. **B**) 2D function.

derivative kernels [Lindeberg, 1998] (Eqn 3.6) allow well-posed comparison of derived measures during multi-scale analyses. $\gamma < 1$ emphasises the results at smaller scales [Krissian et al., 2000], and conversely $\gamma > 1$ emphasises those from larger scales.

$$\frac{\delta}{\delta x}V(x,s) = s^{\gamma}V(x) \otimes \frac{\delta}{\delta x}g(x,s) \tag{3.6}$$

Manipulation of the second derivatives in H can give rise to several measures including the Gaussian curvature and mean curvature, which have been employed for the detection of cortical sulci and blood vessels [Avants and Gee, 2003][Renault et al., 2000]. This kind of multi-scale analysis using H can be envisaged as a probe (Fig 3.6) measuring the contrast between the regions inside and outside the range $[-s,s]$. More complex descriptions of the local shape than simple curvature-based measures can be created by considering the eigensystem of H (Dfn 3.3.1): the eigenvectors υ can be viewed as the principal axes defining an ellipsoid of intensity curvature and the eigenvalues λ have an inverse relationship to the axes' individual lengths[7]. Sorting this eigensystem by the eigenvalues' magnitudes means negative λ_3 describes a bright region – curvature rising to an intensity maximum at the current scale-space level – and positive λ_3 denotes a region darker than its local neighbourhood.

Definition 3.3.1 (Eigensystem of the Hessian) The **eigensystem of a Hessian matrix** H for N-dimensional data can be defined by the equation $H\upsilon - \lambda I\upsilon = \mathbf{0}$.

[7]The properties of the eigensystem can also be illustrated in the orientation and eccentricity of an ellipse (Fig 3.7).

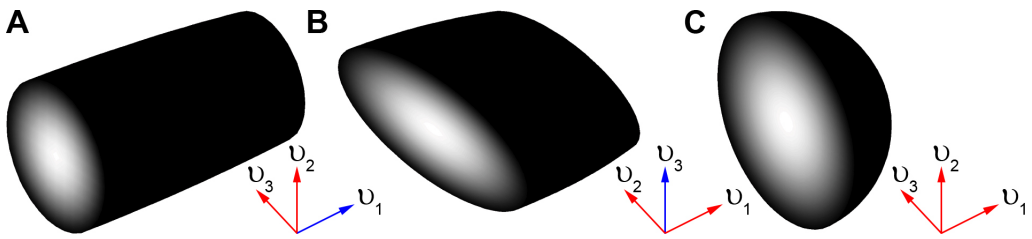

Figure 3.7: **Description of shape primitives**. The intensity profiles of **A**) tube, **B**) plate and **C**) blob shapes are shown in grayscale, where white denotes the most contrasting intensities and black the least. The Hessian eigensystem representations are shown in the lower-right corners. The eigenvectors υ are ranked in ascending order by their eigenvalues λ. Long (low curvature) axes have small eigenvalues and short (high curvature) ones have large eigenvalues. The principal normal direction in υ is highlighted in blue.

The linearly independent eigenvectors υ are the principal directions of curvature of H which form an orthogonal basis set: $\forall i, \upsilon_i \perp \{\upsilon \backslash \upsilon_i\}$. The eigenvalues λ are signed principal curvatures of eigenvectors in υ. The eigensystem is considered ranked by the magnitudes in λ, so υ_1 is the direction of minimal curvature (the longest axis in an ellipsoid representation) and υ_N is the direction of maximal curvature (the shortest axis).

Early works employing this information for detecting of blood vessels in angiographic volumes were extended in [Frangi et al., 1998] in a robust multi-scale approach. Their descriptions of blob-, tube- and plate-shaped structures by structural peculiarities of their eigensystem (Dfn 3.3.2) echo the definitions stated earlier in terms of covariance matrices (Dfn 3.2.1). In all these works bright tubular blood vessels were the structures of interest on a dark background, and so the desired properties of their eigensystems provided the template for constructing a multi-scale tubular structure measure T (Dfn 3.3.3) over a range of scale-space levels $[s_{min}, s_{max}]$. Referring to the objectives stated at the beginning of this section, this structure-specific filter provided sensitivity to the expected scale of the structures as well as some robustness to noise, and the Gaussian derivative filters used to construct H also helped to ensure locally maximal values of T were found along the medial axis of the vessels. Since the local structure orientation could also be derived from υ, then this approach seemed the most suitable for adaptation to the application domain.

Definition 3.3.2 (Primitive shape eigenproperties) A **tube** possesses one direction of minimal intensity curvature υ_1 where $|\lambda_1| < |\lambda_2\lambda_3|$. υ_1 is normal to a cross-section of maximal intensity curvature formed by υ_2 and υ_3, and $|\lambda_2| \approx |\lambda_3|$.

A **plate** possesses one direction of maximal intensity curvature υ_3 where $|\lambda_3| > |\lambda_1\lambda_2|$. υ_3 is normal to a cross-section of minimal intensity curvature formed by υ_1 and υ_2, and $|\lambda_1| \approx |\lambda_2|$.

A **blob** possesses close to umbilical curvature, so $|\lambda_1| \approx |\lambda_2| \approx |\lambda_3|$. The principal directions of curvature υ thus are not uniquely defined.

Definition 3.3.3 (Tubular structure measure) The **tubular structure measure** T contained three subcomponents. First, the **deviation from plate-shape** measure R^A examined the ratio of the largest principal curvatures to discriminate against plates. In the case of a tube or blob, R^A attained its maximum value since $\lambda_2 \approx \lambda_3$ and there were at least two maximal curvature directions. When there was only one (a plate) then $\lambda_2 < \lambda_3$ and the response from R^A was reduced.

$$R^A(\lambda) = \frac{|\lambda_2|}{|\lambda_3|} \tag{3.7}$$

The **deviation from blob-shape** measure R^B described the eccentricity of the second order ellipse which responded well to blobs when $|\lambda_1| \approx \sqrt{|\lambda_2\lambda_3|}$, less to plates where $|\lambda_1| \approx |\lambda_2|$, and poorly to tubes since $|\lambda_1| < \sqrt{|\lambda_2\lambda_3|}$.

$$R^B(\lambda) = \frac{|\lambda_1|}{\sqrt{\lambda_2\lambda_3}} \tag{3.8}$$

The Euclidean norm of λ forms S, the **total magnitude of second-order structure**, which implemented robustness to noise and removed low contrast background since R^A and R^B are gray-level invariant. Tubes should possess two large eigenvalues resulting in high S. Plates will have just one large eigenvalue, leading to a reduced result. The magnitude of the eigensystem is expected to be even smaller for the poorly-contrasting background, producing the lowest values of S. N is the dimensionality of the data.

$$S(\lambda) = \sqrt{\sum_{n \in N} \lambda_n^2} \tag{3.9}$$

Finally, T responds optimally to high R^A and S, and low R^B when the local intensity contrast is bright ($\lambda < 0$) in combination providing sensitivity to contrasting tube-like structures. The maximum response over the eigensystems computed at scale-space levels in the range $[s_{min}, s_{max}]$ is selected on a voxel-by-voxel basis. α, β and c are thresholds controlling the sensitivity to R^A, R^B and S respectively in a manner similar to the variance of a Gaussian function.

$$T(\lambda, s_{min}, s_{max}) = \begin{cases} \max\limits_{s_{min} \leq s \leq s_{max}} (1 - e^{-\frac{R^A(\lambda[s])^2}{2\alpha^2}}) e^{-\frac{R^B(\lambda[s])^2}{2\beta^2}} (1 - e^{-\frac{S(\lambda[s])^2}{2c^2}}), & \text{if } \lambda_3 < 0; \\ 0, & \text{otherwise.} \end{cases} \tag{3.10}$$

In combination, R^A and R^B can determine the local intensity profile as being close to either a contrasting plate, tube or blob. An alternative discriminant is the tensor mode [Criscione et al., 2000], used in [Ennis and Kindlmann, 2006] to improve the interpretability and visualisation of DTI data. The diffusion tensor A was split into isotropic (Eqn 3.12) and anisotropic (Eqn 3.13) components, and then decomposed using three orthogonal variants to characterise the magnitude of isotropy, the magnitude of anisotropy, and the mode of anisotropy (Eqn 3.14). det is the matrix determinant function, norm is the Frobenius norm, λ are the eigenvalues of A, skew is the skewness of a vector and var is the variance of a vector.

$$A = \overline{A} + \widetilde{A} \tag{3.11}$$

$$\overline{A} = \frac{1}{3} trace(A) I \tag{3.12}$$

$$\widetilde{A} = A - \overline{A} \tag{3.13}$$

$$\begin{aligned} \text{mode}(\widetilde{A}) &= 3\sqrt{6}\det\left(\frac{\widetilde{A}}{\text{norm}(\widetilde{A})}\right) \\ &= \frac{\lambda_1\lambda_2\lambda_3}{\text{norm}(\widetilde{A})^3} \\ &= \sqrt{2}\text{skew}(\lambda)\text{var}(\lambda)^{-\frac{3}{2}} \end{aligned} \tag{3.14}$$

The tensor mode has several equivalent forms listed in Eqn 3.14, but the last is the most revealing: the mode is proportional to the variance and skewness of the eigenvalues. The normalising constants were chosen so that the result of the mode is in the range $[-1,+1]$. Three types of anisotropy have specific modes: extreme plate-like shapes occur when $\text{mode}(\widetilde{A}) = 1$, extreme tubular shapes are found when $\text{mode}(\widetilde{A}) = -1$, and blobs naturally have no $\widetilde{A}$ component so $\text{mode}(\widetilde{A}) = 0$. While the decomposition of the three orthogonal variants of A (including the mode) can be useful for visualisation purposes, similar statements about local curvature can be put forward in terms of calculations on eigenvalues (such as ratios in R^A and R^B).

3.4 Estimation of contrasting thin structure presence, shape and orientation in MRI volumes

The multi-scale vessel enhancement filter described in the previous section [Frangi et al., 1998] can highlight thin bright tubular structures. However, it is not responsive to the plate-like profiles of cortical sulci (because of R^A), neither can it detect the dark intensities of blood in MRI (thanks to the restriction on the sign of λ_3 in Eqn 3.3.3) nor does it generate an estimate of the local structure covariance matrix (examined in Section 3.2) needed for the generation of similarly-shaped structure-preserving anisotropic filters. There are several goals relevant to detecting the particular local intensity curvature profiles associated with cortical sulci and blood vessels in dual-echo T_2-PD structural MRI volumes:

- The excellent contrasts of dark blood in PD-weighted data and also of bright CSF structures in T_2-weighted data should be exploited;
- The detection process should not distinguish between tube and plate shapes, since BG- and CSF-containing thin structures may appear in both forms (i.e. the dark plate-like skull has indistinguishable intensity characteristics from blood

vessels) and these local profiles may not be readily distinguishable in MRI volumes with thick slices;

- The performance should decline gracefully in conditions of poor resolution of the thin structures, strong noise, or other factors which may lead to deviation from the ideal intensity profiles as investigated in Section 3.4.4;
- The value of T should not decline as rapidly toward the edges of the relevant structures, which are likely to be affected by PVE;
- The method should respond poorly to blob-like contrasting regions which possess little anisotropy, and also to non-thin BG and CSF structures as well as to noise;
- Maps of thin BG and CSF structures should have a high PPV. This property will allow future processing steps to treat them as reliable priors of their (at least partial) tissue class content inside voxels;
- Locally adaptive filters should be created which are oriented and shaped similarly to the local structure of sulci and vessels, in order to preserve their anisotropy during future filtering processes.

For the modifications proposed in this section the term *principal normal direction* has been coined (Dfn 3.4.1), which augments basic descriptions of local shape as a plate, tube or blob with a single vector (chosen from the principal directions of curvature) which succinctly describes its orientation.

Definition 3.4.1 (Principal normal direction) The **principal normal direction** of a tube is the minimal principal direction of curvature $\upsilon_1 \perp \{\upsilon_2, \upsilon_3\}$ along its length, and for a plate it is the maximal principal direction of curvature $\upsilon_3 \perp \{\upsilon_1, \upsilon_2\}$ across its thin dimension. υ are the ascending-ranked eigenvectors of the Hessian matrix. A blob does not have a uniquely-defined principal normal direction since υ are not uniquely defined.

Five sub-tasks lead to the generation of the adaptive anisotropic filter set. New magnitude of maximal curvature (Section 3.4.1) and deviation from blob-shape (Section 3.4.2) measures lead to the formation of new BG- or CSF-specific thin structure

measures which can thresholded with determination of the locally optimum scale (Section 3.4.3). Their principal normal directions can be extracted (Section 3.4.4) to create estimates of the local structure covariance matrix. To improve robustness to noise and propagate the conditions found at the medial axes of thin structures toward their edges, adaptive smoothing of the Hessian is performed (Section 3.4.5). Procedures for training the expected thin structure scales are examined in Section 3.4.6 and finally a proposal is made for removing unwanted remnants of non-thin contrasting structures (Section 3.4.7).

3.4.1 Magnitude of maximal curvature measure

The total magnitude of second-order structure S (Eqn 3.9) defined in [Frangi et al., 1998] assumed that noisy regions comprising low-contrast, non-vessel structures possessed low principal curvatures (small λ) compared to vessels in MRA, and hence the Euclidean norm of λ would be an effective discriminant. However, in MRI of the brain these "background" structures (such as WM and GM) have a much greater range of intensities, and also a stratification emerges in the Euclidean norm according to the local structure shape (summarised in Dfn 3.2.1): contrasting plate shapes possess just one large eigenvalue, contrasting tubes have two, and all the eigenvalues are similar and thus will be large for contrasting blobs. For the new contrasting structure discrimination measure the determination of local shape is not important (because thin BG and CSF structures can take both shapes, as noted in the introduction to this section) but the equality of response to tube and plate shapes is vital and low-contrast regions should still be rejected.

The new magnitude of maximal curvature measure S_{new} (Dfn 3.4.2) hence eliminates shape dependency from S. Binary sensitivity to either locally bright or dark structures is implemented through a check on the sign of λ_3 using the *contrast* flag, so only dark BG structures will be detected in PD volumes and T_2 volumes will reveal bright CSF, where they respectively possess the best contrast (as shown in Fig 1.11).

Definition 3.4.2 (Magnitude of maximal curvature measure) The **magnitude of maximal curvature** S_{new} responds with zero output if the largest principal curvature possesses the wrong sign (dictated by the flag $contrast \in \{dark, bright\}$),

and otherwise returns the absolute magnitude of the largest principal curvature.

$$S_{new}(\lambda, contrast) = \begin{cases} |\lambda_3|, & \text{if } [(contrast = bright) \wedge (\lambda_3 < 0)] \\ & \vee [(contrast = dark) \wedge (\lambda_3 > 0)] \\ 0, & \text{otherwise.} \end{cases} \quad (3.15)$$

3.4.2 New deviation from blob-shape measure

The magnitude of maximal curvature measure S_{new} (proposed in Dfn 3.4.2) does not discriminate which local shape is present, but since blobs are not desired then the deviation from blob-shape measure R^B (Eqn 3.8) proposed in [Frangi et al., 1998] has been adapted for this purpose. Like S (Eqn 3.9), R^B stratified its response to each shape prototype: the ratio of the smallest eigenvalue to the two largest was smallest for tubes which possessed two large eigenvalues, the result was increased for plates which have one large eigenvalue, and the highest response was induced by blobs where all the eigenvalues are large. In order to combat the distinction between tubes and plates, the dependence on λ_2 has been removed in an updated deviation from blob-shape measure R^B_{new} (Dfn 3.4.3) so that only the relative disparity between the smallest and largest eigenvalues remains, which determines whether a blob shape is present.

Definition 3.4.3 (New deviation from blob-shape) The new **deviation from blob-shape** measure R^B_{new} responds with maximal output to blobs, where $\lambda_3 \approx \lambda_1$. Reduced responses are produced for both tubes and plates, where $\lambda_3 > \lambda_1$.

$$R^B_{new}(\lambda) = \frac{|\lambda_1|}{|\lambda_3|} \quad (3.16)$$

3.4.3 Computation of the thin structure measure

The introduction to this section stated that thin structures containing BG or CSF can take both plate and tube shapes, and so the form of the original tubular structure measure T (Eqn 3.10) proposed in [Frangi et al., 1998] can be simplified by the removal

of the deviation from plate-shape measure R^A (Eqn 3.7) which discriminates between them. The new thin structure measure T_{new} (Dfn 3.4.4) incorporates the updated deviation from blob-shape measure R^B_{new} (Dfn 3.4.3) in place of R^B (Eqn 3.8) as well as the magnitude of maximal curvature measure S_{new} (Dfn 3.4.2) which replaces S (Eqn 3.9). These measures in combination cause the optimal response of T_{new} to be generated by contrasting plate or tube shapes which are either bright in T_2 volumes (for CSF) or dark in PD volumes (for BG). The outputs of R^B_{new} and S_{new} are non-negative values which are modified by negative exponential functions in T_{new} to reduce sensitivity to noise and the "background" (such as WM and GM), with their respective weights determined by their normalising factors β_{new} and c_{new}, similar to the original T.

Definition 3.4.4 (Thin structure measure) The **thin structure measures** T^{CSF}_{new} and T^{BG}_{new} produce high output when the local shape is not blob-like (the R^B_{new} term is low), the magnitude of maximal curvature is large (high S_{new} term) and the contrast is appropriate (*bright* or *dark*). max is the maximum operator.

$$T^{CSF}_{new}(\lambda, \beta_{new}, c_{new}) = \left(e^{-\frac{R^B_{new}(\lambda)^2}{\beta_{new}}} \right) \left(1 - e^{-\frac{\left(\frac{S_{new}(\lambda, bright)}{\max S_{new}(\lambda, bright)} \right)^2}{c_{new}}} \right) \tag{3.17}$$

$$T^{BG}_{new}(\lambda, \beta_{new}, c_{new}) = \left(e^{-\frac{R^B_{new}(\lambda)^2}{\beta_{new}}} \right) \left(1 - e^{-\frac{\left(\frac{S_{new}(\lambda, dark)}{\max S_{new}(\lambda, dark)} \right)^2}{c_{new}}} \right) \tag{3.18}$$

The scale preference exponent γ (required in Eqn 3.6 for gradient calculations forming the Hessian) was set to 1 since no preference for whether smaller or larger scales were desired – instead, a tight range of expected thin structure scales $[s_{min}, s_{max}]$ was considered (with the training of these parameters considered in Section 3.4.6). Thresholding the maximum response for T over scales $[s_{min}, s_{max}]$ using restrictive class-specific thresholds $0 \leq \{\pi^{CSF}, \pi^{BG}\} \leq 1$ produced a binary map B (Dfn 3.4.5) with a high PPV, which then ensured B could be employed as both a tissue prior and

also as a selector of where anisotropic filtering should be applied.

Definition 3.4.5 (Scale-selective binary thin structure map) The **thin structure binary maps** B^{CSF} and B^{BG} return a two-element vector. The first element holds the scale-space level *ms* leading to the maximum response of T^{CSF}_{new} or T^{BG}_{new} at each voxel. The second is a flag describing whether the value of T^{CSF}_{new} or T^{BG}_{new} calculated at *ms* has at least reached the appropriate threshold π^{CSF} or π^{BG}. arg max is the argument maximum function which chooses the scale s which maximises T^{new} on a per-voxel basis.

$$\begin{aligned} B^{CSF}(\lambda, s_{min}, s_{max}, \pi^{CSF}) &= \left[ms, T^{CSF}_{new}(\lambda[ms]) \geq \pi^{CSF}\right] \text{ s.t.} \\ & ms = \arg \max_{s_{min} \leq s \leq s_{max}} T^{CSF}_{new}(\lambda[s]) \end{aligned} \tag{3.19}$$

$$\begin{aligned} B^{BG}(\lambda, s_{min}, s_{max}, \pi^{BG}) &= \left[ms, T^{BG}_{new}(\lambda[ms]) \geq \pi^{BG}\right] \text{ s.t.} \\ & ms = \arg \max_{s_{min} \leq s \leq s_{max}} T^{BG}_{new}(\lambda[s]) \end{aligned} \tag{3.20}$$

3.4.4 Adaptive shaped and oriented filtering

Filters which are shaped and oriented similarly to the local structure covariance matrices have been shown to help preserve their anisotropy in smoothing operations (as investigated in Section 3.2). In the Discussion a functional relationship between the covariance of these filters and the principal curvatures λ of the local structure has been suggested, but for simplicity and robustness, the variance variables *large* and *small* used in Dfn 3.2.1 are considered constant over the volume in this section. The same motivations lead to assumptions that bright CSF structures will be best modelled with a filter possessing a plate-like shape and dark BG structures by tube-like ones. This decision has been made since in clinical MRI brain volumes the deviation from plate-shape measure R^A (Dfn 3.7) cannot be thresholded to discriminate well between plate and tube shapes (i.e. the shapes of tubular blood vessels from the plate-like skull in Fig 3.9F), indicating that there may be problems in accurately calculating the smallest components of the eigensystem of H. The calculation of the eigensystem is affected

both by noise in the intensities which will modify the smallest (low-frequency) components of the eigensystem relatively more than the others, and also by the poor resolution of the thin structures – a problem exacerbated by high slice thickness (which is double the axial-plane voxel dimensions in all the datasets examined in this thesis except for BrainWeb [Cocosco et al., 1997]). The latter issue is especially apparent for thin structures running in the axial-plane direction, since then PVE reduces their in-plane contrast and also their through-plane intensity profile.

Concerning the application of anisotropic filters, their covariances Σ must be made relevant to the estimated local orientation of thin structures. The convention used in Section 3.2 was for the principal normal direction (Dfn 3.4.1) to occupy the top-left element of each prototype shape covariance matrix and so the required rotation angle can be calculated as $acos(u \cdot d)$ between a unit vector pointing along this first axis ($u = [1,0,0]$) and the estimated local principal normal direction d, on the plane $u * d$ which is orthogonal to these vectors. The MATLAB `makehgtform` function with the `axisrotate` option, which allows for rotation around an arbitrary axis, was used to create the transformation matrix R for Σ.

Storing a filter for each thin structure voxel is an unrealistic prospect due to high potential memory usage: a typical volume with $256 \times 256 \times 128$ voxels, each able to possess a $5 \times 5 \times 5$ filter stored as MATLAB single precision numbers, would occupy a contiguous 4GB block of memory. However, *large* and *small* are constants and the filters are symmetric Gaussian functions, and so half of the rotation space can be discretised into angles θ_{XY} and θ_{YZ}, and exemplar tube- and plate-shaped filters for each orientation and shape combination can be generated to form a filter bank ϕ with considerable storage cost savings. The granularity of the rotation space does not need to be especially coarse since the memory required to store individual filters is small. Each thin structure voxel can then reference the filter in ϕ with the closest principal normal direction to its own (or to its reflection), selected via the lowest Euclidean distance between the vectors. Only one half of the space needs to be covered (as pictured in Fig 3.8) since the oriented filters for principal normal directions d and $-d$ are identical.

3.4.5 Noise reduction and smoothing of the eigensystem

The eigensystem of the Hessian H has been used in Sections 3.4.1-3.4.3 to characterise the local structural peculiarities of contrasting plates and tubes. H has been made

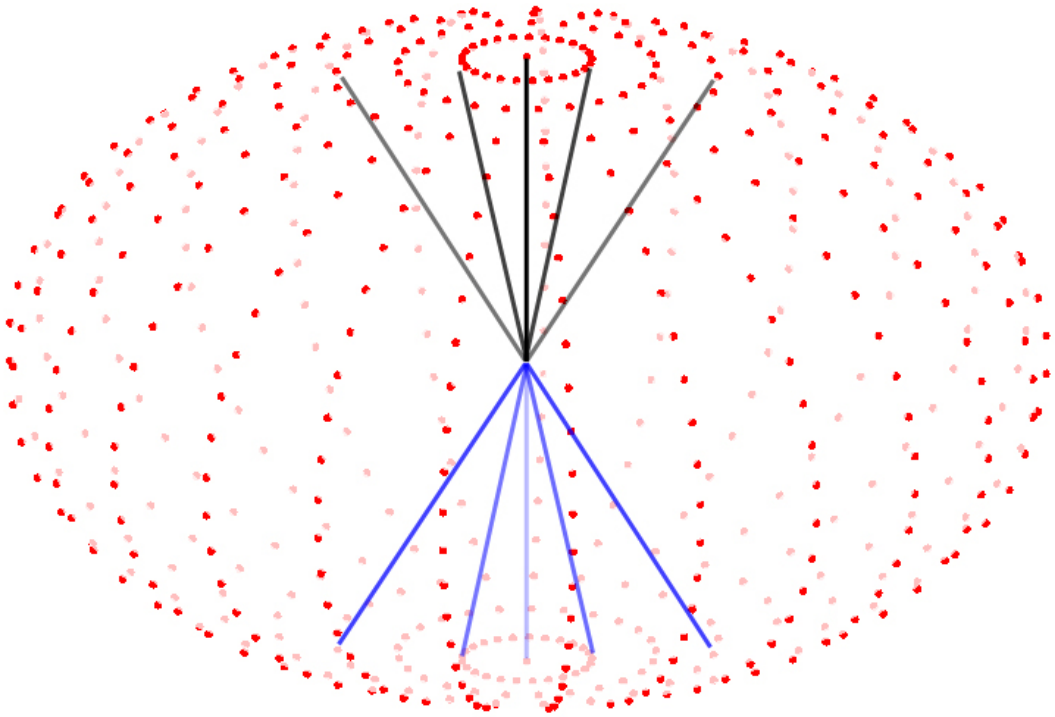

Figure 3.8: **One half of the rotation space is sufficient to cover all filter orientations**. 3D globe showing a discretised rotation space (red dots), with one half of the rotation space mapped (solid black lines) and the other half covered by reflection (solid blue lines).

suitable for multi-scale analysis by its formation using the scale-normalised Gaussian derivative formulations [Lindeberg, 1993] (Eqn 3.6) though high-frequency noise, which occurs on similar scale-space levels to which thin structures are best detected, may affect its calculation with relatively larger effects on the small (lower-frequency) components of its eigensystem as noted in Section 3.4.4. Poor resolution of blood vessels and cortical sulci also means that a large proportion of voxels in which they are present will contain tissue boundaries and thus some (unknown) degree of PVE, causing their intensity contrast to be reduced; at least λ_3 will decrease as a result, affecting S_{new} (Dfn 3.4.2) and R^B_{new} (Dfn 3.4.3) and in turn T_{new} (Dfn 3.4.4) and B (Dfn 3.4.5).

In order to reduce the effects of noise and propagate the optimal conditions for thin structure detection found at the medial axes of thin structures towards their boundaries, then an adaptive process must ensure that noise and "background" tissue (not thin BG nor CSF structures) contribute little to this smoothing and that thin structure regions provide the greatest contribution. Vessel-enhancing diffusion [Manniesing and Niessen, 2005] has extended vessel enhancement filtering [Frangi et al., 1998] to combat noise by smoothing vessel intensities anisotropically along their lengths and more isotropically elsewhere, though it uses the noise- and PVE-affected eigensystem directly to drive diffusion. Smoothing intensities only indirectly results in smoothing components of the eigensystem, and furthermore, smoothing towards the boundaries of thin structures is problematic when performed at the original resolution since the decay of the thin structure measure T_{new} (Dfn 3.4.4) towards the PVE-affected boundary voxels of the structures is rapid, so this approach has not been used. The decay is observed in the poor extension of T-thresholded voxels to sulcal boundaries in Fig 3.9C, while high specificity of the thresholded map is maintained.

In the following proposal, H has been adaptively smoothed rather than the intensities, T_{new} or the eigensystem. Each element of H has a precise meaning as a particular second derivative at each voxel, and because H forms a symmetric matrix at each voxel, then simple definitions of local orientation coherence can be formed. Orientation coherence can then supplement measures of local contrast to allow smoothing to be performed differently along well-defined, small-scale contrasting structures than elsewhere. Orientation coherence of the eigenvectors is more difficult to calculate than that of H (since their reflections must also be considered) but because smoothing of H would also manifest in its eigensystem, and subsequently in the computation of both T_{new} and the estimate of the local structure covariance matrix, then this option was preferred.

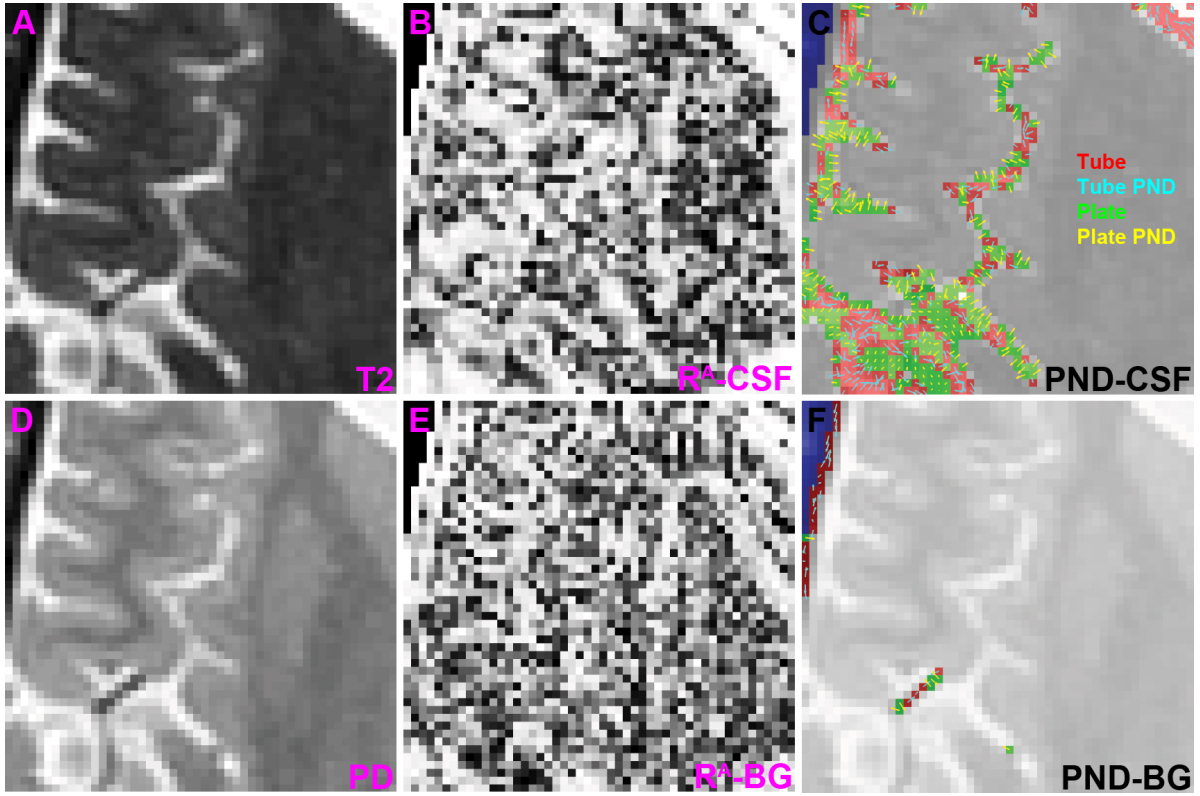

Figure 3.9: **Poor recovery of the extent of thin structures without smoothing of the Hessian, and poor discrimination between plate and tube shapes using** R^A. Data from ROI 5 in the LBC1936 dataset are shown. Detection was performed over scales $[0.7, 0.8, 0.9]$ for CSF and $[0.5, 0.55, 0.6]$ for BG. R^A: Plate-tube shape discriminant (Dfn 3.7). **PND**: Principal normal directions (Dfn 3.4.1) shown by the coloured vector field for thresholded CSF structures (where $T^{CSF} > 0.345$) and thresholded BG structures ($T^{BG} > 0.045$). Voxels highlighted in red are proposed to be tube-like (thresholded by $R^A \geq 0.85$); voxels in green are proposed to be plate-like ($R^A < 0.85$).

A related area of research is the smoothing of water motility derivatives from DTI data [Zhang and Hancock, 2006] for which normalised convolution [Westin and Knutsson, 2003] (Dfn 3.4.6) has been employed. This method has been chosen to be modified for the task at hand due to its incorporation of necessary local orientation and magnitude similarity terms, and since it completes in a single pass with no termination conditions required then the possible extent of smoothing is precisely known (as opposed to vessel enhancing diffusion where termination conditions are required to stop iterations).

Definition 3.4.6 (Normalised convolution) **Normalised convolution** uses a least squares approach, weighted by a matrix W, to minimise the distance between a noisy signal F and its reconstruction from basis functions B parameterised by Θ.

$$\begin{aligned} \|W(B\Theta - F)\| &= 0 \\ (WB)^T WB\Theta &= (WB)^T WF \\ B\Theta = F_{new}(W,B,F) &= ((WB)^T W)^{-1}(WB)^T WF \end{aligned} \tag{3.21}$$

If only a single basis function is used, then effectively it can be replaced with the identity matrix ($B = I$) and the reconstruction is simplified.

$$F_{new}(W,F) = (W^T W)^{-1} W^T WF \tag{3.22}$$

The weight $W^T W = W_a(\Sigma)W_c(H, H_{nhood}, \sigma_m, \alpha)$ was the product of two factors. Firstly, the applicability function W_a implemented the decreasing power of neighbourhood context with increasing distance from a central position during smoothing. It was made spatially-adaptive, to anisotropically smooth most across the thinnest dimensions of thin structures: tube-shaped filters can be aligned along the thin dimension of plate-like CSF structures, and conversely the normals of plate-shaped filters can be aligned along the long dimension of tube-like BG structures. Secondly, the local data certainty matrix W_c (Dfn 3.4.7) guided smoothing by combining several measures: a binary brain parenchyma mask c_b which eliminates non-brain data from consideration, local similarities of contrast in the degree of contrasting structure c_n, and also local similarity of Hessian orientation in the local angular similarity c_a. This last component of W_c

was left identical to that presented in [Westin and Knutsson, 2003]. It could identify the boundaries of structures, where the local angular similarity of H is expected to be minimal; smoothing across these boundaries was not desired. c_n has been formed by isotropically-smoothing S_{new} (derived from the unsmoothed H), which takes locally maximal values at the medial axes of thin contrasting structures and helps to reduce contributions from outside them. R^B_{new} was not used in c_n due to the noisiness of the smallest components of the eigensystem derived from the unsmoothed Hessian (shown in Fig 3.13B and G).

In combination, W_a and W_c will cause the conditions leading to high T_{new} (high c_n) at the medial axes of thin contrasting structures (masked with c_b) to be propagated towards their edges (by adaptively smoothing anisotropically with W_a) but proceed no further (with the edges found using c_a). Ideally, the smoothing will allow the thresholds producing the binary maps B to be raised – increasing the specificity of the results – and improve the uniformity of the computed principal normal directions.

Definition 3.4.7 (Thin structure certainty) The **thin structure certainty** W_c is formed by the product of several normalised variables. The certainty is high when brain parenchyma is present ($c_b = 1$), the thin structure measure (c_n) returns a high value, and local angular similarity (c_a) is high.

$$W_c(H, H_{nhood}, \Omega, \lambda, \alpha) = c_b * c_n(H, \Omega, \lambda) * c_a(H, H_{nhood}, \alpha) \tag{3.23}$$

The brain parenchyma mask $c_b \in \{0, 1\}$ is a binary variable whereas the others are continuous. The degree of contrasting structure c_n is derived from the magnitude of maximal curvature S_{new} (Dfn 3.4.2), and is high when the local intensity contrast is strong – dark for vessels or bright for sulci. Since S_{new} displays very low values for "background" tissue, $\Omega > 0$ raises the value of S_{new} from its low base so that smoothing will properly occur among non-thin structure voxels; otherwise the contribution from thin structure voxels at the edge of W_a's filter coverage will dominate. max is the maximum operator.

$$c_n(H, \Omega, \lambda) = \frac{S_{new}(\lambda) + \Omega . \max\left[S_{new}(\lambda)\right]}{(1 + \Omega)\max\left[S_{new}(\lambda)\right]} \text{ s.t. } \lambda = \text{eigenvalues}(H) \tag{3.24}$$

The local angular similarity c_a between tensor H and its neighbourhood average H_{nhood} is defined by the inner product of their normalised forms. The exponent α controls the extremity of the output.

$$c_a(H, H_{nhood}, \alpha) = \left\langle \frac{H}{\|H\|}, \frac{H_{nhood}}{\|H_{nhood}\|} \right\rangle^{\alpha} \tag{3.25}$$

H_{nhood} is the neighbourhood average of H required to form c_a, which can also be calculated using normalised convolution, but with a simpler weight $W^T W = W_i(\sigma) W_l(H, \Omega, \lambda)$. An isotropic Gaussian smoothing filter W_i was substituted for W_a so that the calculation of c_a would not be affected by the different anisotropic filter shapes, and a limited thin structure certainty matrix W_l (Dfn 3.4.8) contained the sub-terms of W_c that were not dependent on H_{nhood}, in place of W_c.

Definition 3.4.8 (Limited thin structure certainty) The **limited thin structure certainty** W_l is formed by the product of the binary brain parenchyma mask $c_b \in \{0,1\}$ and the continuous normalised degree of thin structure c_n (both defined in Dfn 3.4.7). The limited certainty is high when brain parenchyma is present ($c_b = 1$) and c_n returns a high value.

$$W_l(H, \Omega, \lambda) = c_b * c_n(H, \Omega, \lambda) \tag{3.26}$$

To allow sub-voxel accuracy in smoothing to thin structure boundaries, PVE must be considered since it is a major factor in the parent voxels. Since it has been stated that a scale-space representation cannot perform better than its discrete realisation [Weickert, 1996], and in [Cachia et al., 2003] it was argued that derivative values taken from most discrete grids are not sufficiently accurate for tracking vessels along their medial axes at typical clinical resolutions, then one course of action is to artificially increase the resolution of the intensity data using an interpolation process to provide estimates of the sub-voxel values. While linear interpolation between sample points will blur well-discretised boundaries, this issue is alleviated to some degree by higher-order interpolants. Cubic interpolation provides a good compromise between ringing artifacts, edge blurring and aliasing [Lehmann et al., 1999] and this method was applied chiefly

because much steeper changes between sub-voxel intensities (resulting in better effective reduction of PVE) are possible than with linear interpolation.

3.4.6 Selecting relevant scale-space levels

The efficiency and accuracy of the thin structure measure T_{new} (Dfn 3.4.4) is dependent upon investigating only a narrow range of scale-space levels, denoted $[s_{min}, s_{max}]$. No formal procedure for selecting $[s_{min}, s_{max}]$ was detailed in [Frangi et al., 1998], so a practical manual procedure has been developed (Proc 3.4.1) to quantify thin structure thicknesses based on just a few regions representing a population of brain volumes.

Procedure 3.4.1 (Thin structure thickness measurement procedure) The following protocol should be followed to obtain the thicknesses of blood vessels or cortical sulci:

- Samples of the particular thin structures should be selected where the volume slice cuts along their medial axes – so cortical sulci proceed in a perpendicular direction, and blood vessels run parallel to the slice;
- Voxels with any content of the particular thin structures (including partial volumes) on the slice should be flagged. The start and end voxels of each thin structure must also be marked. The ROI labelling tool (described in Section 2.4) was adapted for this purpose;
- The shortest path between each start and end voxel pair, travelling through the flagged voxels, is calculated for each thin structure;
- The thickness of each thin structure is determined by dividing half the number of their flagged voxels (including the start and end voxels) by their path length.

The relationship between thin structure thicknesses and the scale-space levels that give the optimum value of T_{new} at the medial axes, for samples taken from the LBC1936 dataset, is investigated in Fig 3.10C. A trend line fitted to this scatter plot using the least squares distance revealed the ratio of the thickness to the optimal scale-space level to be approximately 1.4:1, which roughly held when the blood vessels and cortical sulci

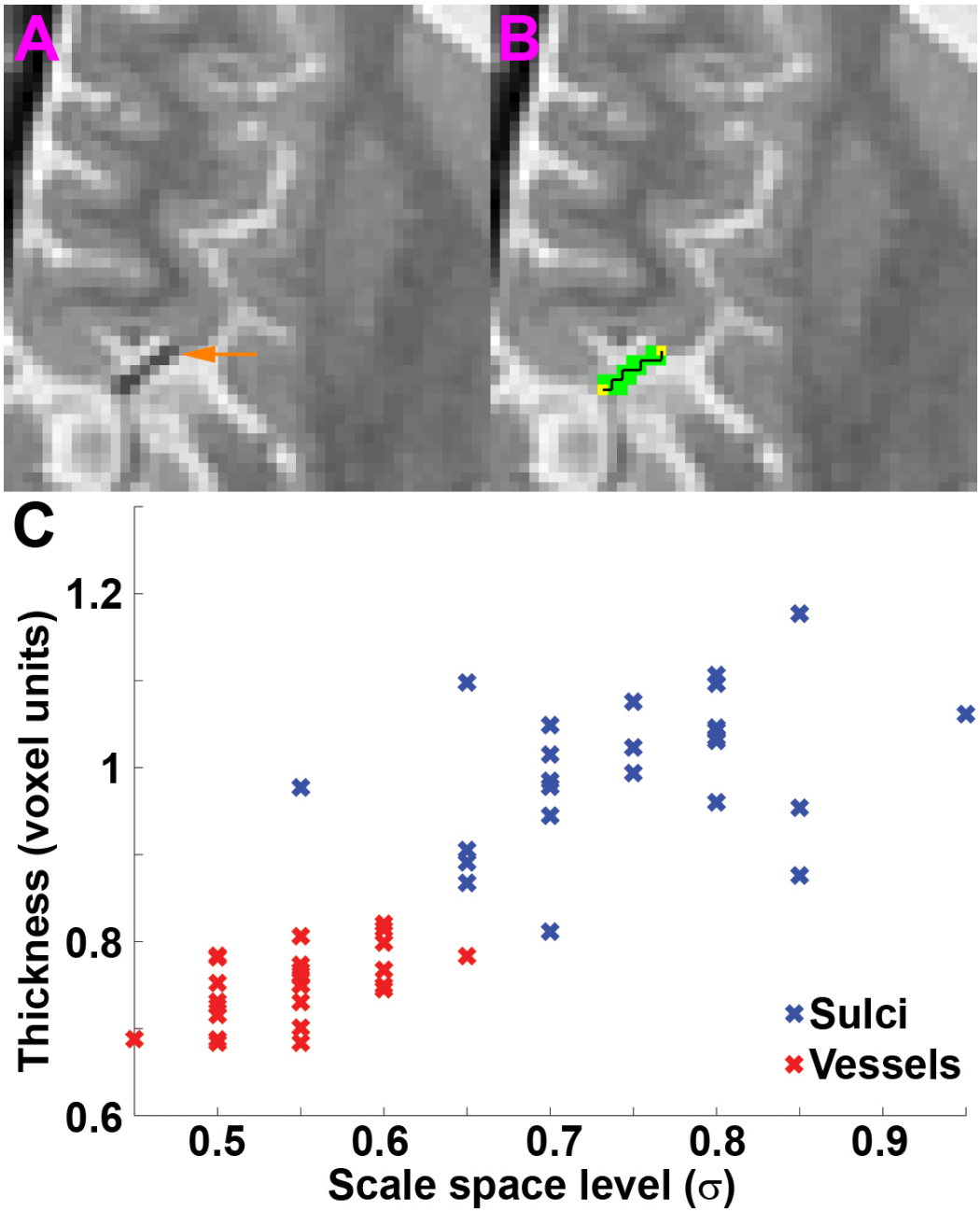

Figure 3.10: **Measuring thin structure thicknesses using Procedure 3.4.1**. **A**) PD data from ROI 5 in the LBC1936 dataset, with a vessel (marked with an orange arrow) running parallel to the slice. **B**) Voxels in **A** containing any part of the vessel are labelled in green; the start and end voxels are labelled in yellow. The shortest path (between voxel centres) between the start and end voxels is marked with a black line. **C**) Scatter-plot of the scale-space level giving the highest appropriate mean T_{new} response against the estimated radial thickness of the structure using this procedure. 50 samples of thin structures in the young brain dataset were chosen; 25 from blood vessels (red crosses), and 25 from cortical sulci (blue crosses).

were investigated separately. Since the thicker regions of CSF do not respond as poorly to isotropic filtering and are generally more easily identified during segmentation, the interest is in the thin (or thin parts of) sulci which makes this method for determining $[s_{min}, s_{max}]$ robust to even severely degenerated brains with wide sulci.

3.4.7 Proposal for removal of larger contrasting structures

The thin structure detection methods outlined in the previous subsections also highlight the borders of large contrasting regions such as the ventricles and the largest cortical sulci. Closer examination of T_{new} (Dfn 3.4.4) maps revealed that the high contrast of the relatively-pure tissue voxels lead to high S_{new} and R^B_{new} (as observed in Figs 3.12B and 3.13B) caused by high curvature to the darker surrounding tissue (greater in magnitude than for PVE-affected smaller structures), despite their lack of intensity contrast in one orientation. While these regions are true positives for the tissue content prior maps, they are not desired for small-scale-specific processing, and will experience an increased sensitivity to noise if processed as such.

Existing methods used to specifically identify the ventricles – the quintessential large contrasting structures in structural T_2-weighted brain MRI – usually require expert assistance: the refinement of borders [Worth et al., 1998], seed placement [Dalton et al., 2002] for region growing methods, or the enforcement of topological and morphological rules [Schnack et al., 2001] all need manual tuning. Given that the borders of larger regions tend to be highlighted by the thin structure detection method, then a simple test for whether these edges enclose a region of pure tissue could conceivably involve a flood fill-style operation; however, the regions are not guaranteed to be fully enclosed by flagged sub-voxels. Furthermore, this problem cannot be reliably solved by calculating the Hessian at larger scales due to variability in size (especially in aged brains) and also since tightly-packed contrasting structures – such as the sulci in the cerebellum (pictured in ROI 4, Fig 2.2) – may present false positives.

Instead, an automatic and pragmatic approach is detailed in Proc 3.4.2 (and pictured in Fig 3.11) which uses intensity-based segmentation with some notion of neighbourhood content homogeneity to locate bright contrasting pure tissue voxels congregating in large regions. A combination of morphological operations applied to the thresholded class maps generated by segmentation can then reliably identify these regions.

Procedure 3.4.2 (Large contrasting region identification) Masks to remove large contrasting regions of tissue from the thin structure maps can be created as follows:

1. The expected range of thin structure scales $[s_{min}, s_{max}]$ is obtained using Proc 3.4.1;
2. A simple and robust segmentation algorithm, which includes homogeneity constraints based on neighbouring elements' intensities, is used to produce membership values of the four main tissue classes;
3. A membership threshold $0 < \pi^{large} \leq 1$ on the target class produces a binary mask η;
4. η is eroded using a spherical structuring element to remove thin structures with thicknesses less than or equal to $\text{ceil}(s_{max}) + X$, where $X > 0$ is a buffer thickness;
5. η is then dilated using another spherical structuring element by $\text{ceil}(s_{max}) + Y$, where $Y \geq X$, to at least restore the volume of the larger regions and thus complete the formation of the large contrasting region map.

The values of parameters X and Y used in the morphological operations detailed in steps 4 and 5 of Proc 3.4.2 should be set generously, so that small regions are fully removed and the boundaries of the larger regions are fully recovered. Once the larger region mask η has been produced, it can be subtracted from the original thin structure maps B (Dfn 3.4.5).

3.4.8 Processing steps and algorithm summaries

Procedure 3.4.3 (Thin structure detection) Thin BG and CSF anatomical structures can be detected from dual-echo T_2-PD volumes by using the following procedure:

1. The volumes must first undergo skull stripping as described in Section 2.2;

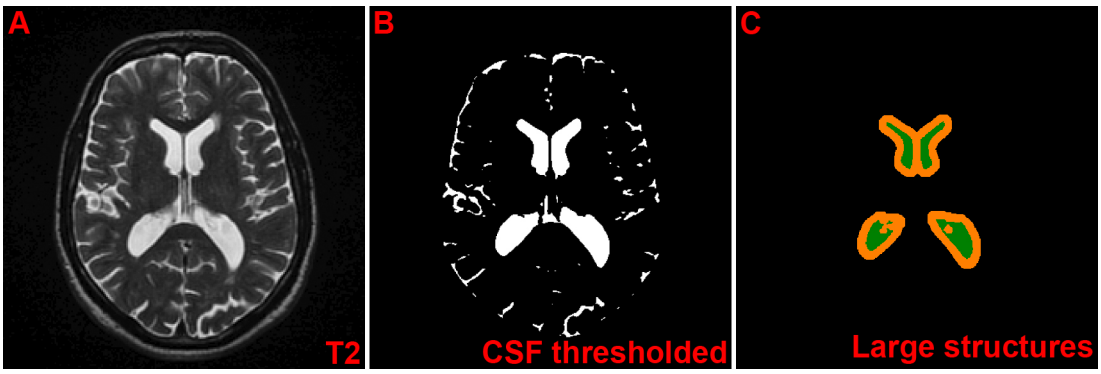

Figure 3.11: **Simple method for identification of large regions of contrasting tissue**. **A**) T_2 axial slice on which ROI 5 is located. **B**) Post-thresholding of memberships of the CSF class (determined by the FCM clustering algorithm described in Section 5.3.2) with $\pi^{large} = 0.5$. **C**) Erosion by a spherical structuring element with radius $ceil(s_{max})+1 = 2$ voxels to remove the thin structures (with the remaining portion of the ventricles shown in green) and then dilation with another spherical structuring element with radius $ceil(s_{max})+3 = 4$ voxels to cover the border of the larger structures (further area restored shown in orange).

2. The ranges of scales $[s_{min}, s_{max}]$ to consider for blood vessels and cortical sulci are determined separately, using Proc 3.4.1;

3. The volumes are upsized by ratio L via cubic interpolation, and the range of scales is also increased by L;

4. H is calculated using scale-normalised gradients (Eqn 3.6) on the upscaled volumes, and then adaptively smoothed as described in Section 3.4.5;

5. The eigensystem of H is calculated over $[s_{min}, s_{max}]$, and eigenvectors υ for each scale are sorted by their eigenvalues λ;

6. The scale-selective thin structure map T_{new} (Dfn 3.4.4) is calculated, thresholded and stored in B (Dfn 3.4.5) along with the optimum scale-space level at each voxel;

7. Large contrasting structures such as the ventricles are removed from B, using the method described in Section 3.4.7.

Procedure 3.4.4 (Filter production) Bespoke shaped and oriented anisotropic filters can be produced to aid the preservation of thin structures in future neighbourhood filtering operations by using the following procedure:

1. Values of *small* and *large* are retrieved to determine the ratio between the filter variances in each dimension (listed in Dfn 3.2.1);
2. A set of exemplar filters are generated at a number of orientations in half of the orientation space for both plate and tube shapes, as described in Section 3.4.4;
3. The filter shape is set to a tube for sub-voxels in B^{BG} and to a plate for sub-voxels in B^{CSF};
4. The principal normal direction of each thin structure voxel flagged in B is chosen from υ at the optimum local scale-space level, with respect to the constraint expressed in step 3;
5. The filter index for which the principal normal direction of the filter (or its reflection) most closely matches that of each sub-voxel is recorded in ϕ (as described in Section 3.4.4).

3.5 Demonstrations

This section investigates the methods described in Section 3.4 in the context of their accuracy of finding thin BG and CSF structures in the expert-labelled LBC1936 dataset (supplemented by some regions from the young brain dataset), and also with respect to their performance alongside parameter changes. Some of the qualitative assessment was performed with the assistance of Dr. Farrall to ensure that there was no gross misunderstanding of brain anatomy. Unless otherwise stated, the following investigations use the following parameters:

- The range of scale-space levels investigated was $[0.7, 0.8, 0.9]$ for CSF structures and $[0.5, 0.55, 0.6]$ for BG structures (using Proc 3.4.1 to find $[s_{min}, s_{max}]$ for the LBC1936 dataset);

- The scale-normalisation exponent $\gamma = 1$ (no scale preference) was used in the calculation of the scale-normalised gradients (Eqn 3.6) for the Hessian H;

- Normalising values $\beta_{new} = 0.5$ and $c_{new} = 0.5$ were applied to R^B_{new} (Dfn 3.4.3) and S_{new} (Dfn 3.4.2) as suggested in [Frangi et al., 1998]. These values appeared to provide reasonable contrast after investigation of their parameters' effects in Section 3.5.1;

- Thresholds $\pi^{CSF} = 0.345$ and $\pi^{BG} = 0.045$ for the thin structure measure T_{new} (Dfn 3.4.4) were selected after training on ROI 5 to achieve close to a 90% PPV;

- $L = [2,2,2]$ was set as the upscaling ratio for the Hessian smoothing process. A higher ratio was not set due to processing time and memory constraints;

- The angular similarity c_a (defined as part of the Hessian smoothing process in Section 3.4.5) used the exponent $\alpha = 2$. This was a conservative value, chosen to create fairly strong minima surrounding thin structures in the Hessian certainty W_c (Dfn 3.4.7) and ensure there was little contamination between thin- and non-thin structure regions;

- Masks for regions surrounding and including large CSF areas (such as the ventricles) produced in Section 3.4.7 supplemented the parenchyma masks produced in Section 2.2, so that these regions were excluded from quantitative analyses as they did not contain thin structures. The membership threshold $\pi^{large} = 0.5$ on the CSF class, and morphological thickness parameters $X = 2$ and $Y = 3$ were used;

- Shaped and oriented volume intensity smoothing, and also Hessian smoothing through the applicability function W_a detailed in Section 3.4.5, were performed in thin structure regions using anisotropic Gaussian filters with 8 different orientations and a ratio between variance parameters *large* and *small* (used in Dfn 3.2.1) of 2:1. In other regions an isotropic Gaussian filter was applied with standard deviation $\sigma = L$.

ROI 5 (with its anatomy labelled in Fig 2.2) has been studied in the initial demonstrations and remains a benchmark in future chapters' investigations due to its inclusion of a range of anatomical structures including blood vessels, cortical sulci, convoluted temporal cortex, deep GM nuclei, the meninges, the skull, the edge of the lateral ventricle and thin regions of WM. For qualitative observations made on the plotted principal

normal directions, it should be noted that when observing the plotted principal normal directions then the reflection of these vectors is also considered in assessments of uniformity and correctness. Both the ground truth produced by the expert and the binary thin structure maps produced by thin structure detection can be viewed for ROIs 1-6 in Figs 5.10-5.15, and a little additional commentary is made in Section 5.4.6.

In the following quantitative analyses it should be borne in mind that Hessian smoothing is performed at an upscaled resolution, and comparison against the expert labelling must be performed at the original resolution at which it was acquired. Consequently the results where Hessian smoothing was performed needed to be downsampled, through the mean T_{new} values of L-sized blocks of voxels. Since the thresholds on T_{new} had been set at the upscaled level (for use in future chapters) where the contrast of T_{new} was more stark, but the same ones were applied at the downsampled level, then the sensitivity of the downsampled results was reduced as the contrast was lower (due to PVE). Furthermore, the localisation of thin structure boundaries will likely be a little different when processing at a sub-voxel scale because of the Hessian smoothing performed as well as due to working with cubic-interpolated upscaled data. Blood vessels tend to be the most affected by these issues as they are generally thinner than sulci in their thinnest axis, and also because they possess two thin dimensions (as tubular structures) as opposed to one (for plate-like sulci).

3.5.1 Performance of shape deviation and brightness measures

The normalising parameters c_{new} and β_{new} of the new magnitude of maximal curvature measure S_{new} (Dfn 3.4.2) and the new deviation of blob-shape measure R^B_{new} (Dfn 3.4.3) determine how extreme the outputs of these negative exponential functions should be. In [Frangi et al., 1998] it was recommended that equivalent parameters c and β for S (Eqn 3.9) and R^B (Eqn 3.8) both be set to 0.5 to provide robustness to noise and poor local shape definition, and so these values form the starting point for further investigations. In Figs 3.12 and 3.13 one parameter was held constant while the other was varied in order to assess the impact on the agreement of the thresholded thin structure map B (Dfn 3.4.5) with non-zero fractions of the appropriate tissue type in the expert labelling.

In Fig 3.12 β_{new} for R^B_{new} was held constant at 0.5, while c_{new} for S_{new} was assigned extreme values 0.15 and 5. Increasing c_{new} led to increased contrast of T_{new}, and vice versa, though the accuracy that could be obtained through thresholding was very sim-

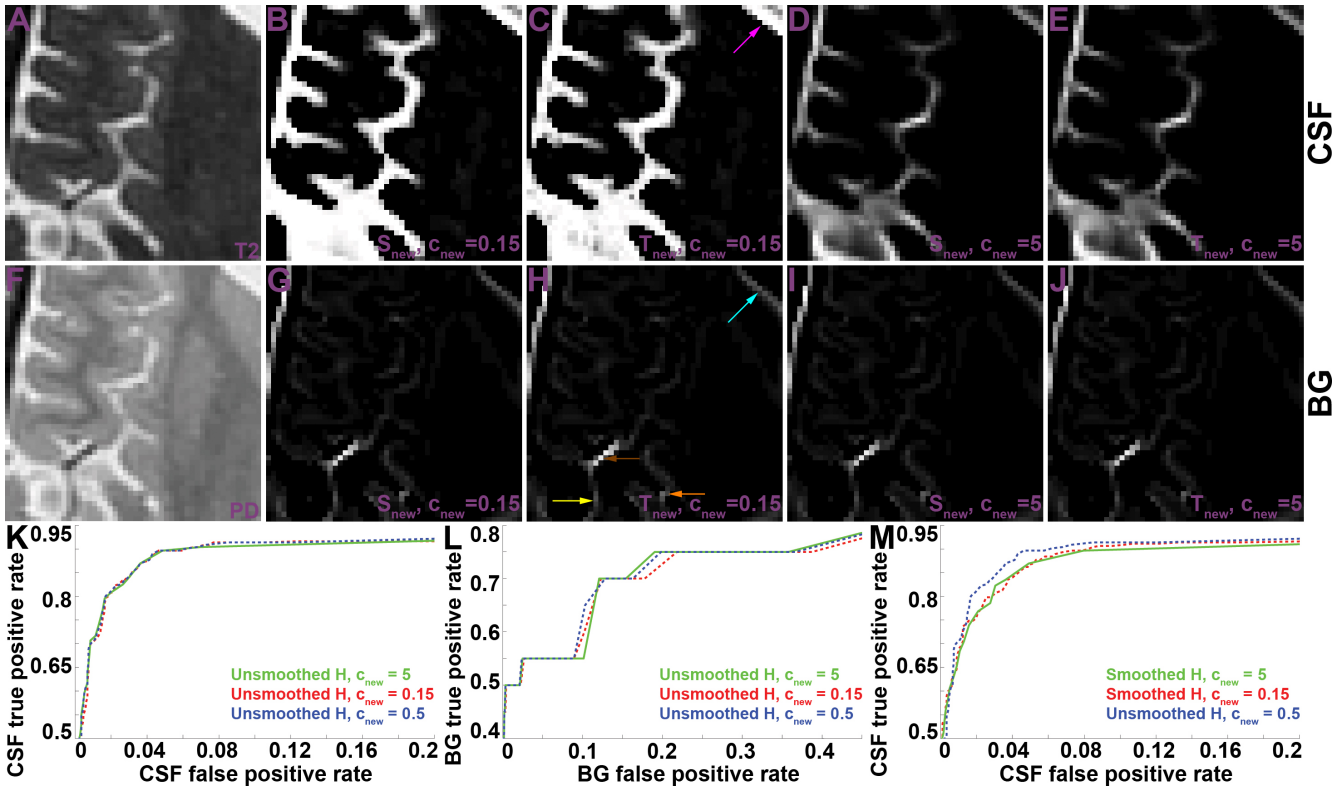

Figure 3.12: **Performance of magnitude of maximal curvature measure** S_{new}. **A** & **F**) T_2 and PD data for ROI 5. **B** & **G**) Value of the S_{new}-containing term of T_{new} (Dfn 3.4.4) for **A** and **F** with $c_{new} = 0.15$. **C** & **H**) T_{new} for **A** and **F** with $c_{new} = 0.15$. **D** & **I**) Value of the S_{new}-containing term of T_{new} for **A** and **F** with $c_{new} = 5$. **E** & **J**) T_{new} for **A** and **F** with $c_{new} = 5$. **K** & **L**) ROC curves for **C**, **E** and **H**, **J** respectively. No Hessian smoothing was performed for panels B–E and G–J, and brighter grayscale values indicate greater values of the relevant measure. Coloured arrows are referenced in Section 3.5.1. **unsmoothed**: no Hessian smoothing performed.

ilar (Fig 3.12K-L). This likely occurred since it can be observed that T_{new} (panels C, E, H and J) is very similar to S_{new} (panels B, D, G and I) in both cases, meaning that the contribution of R^B_{new} was minor. At higher false positive rates (Fig 3.12M) then smoothing of the Hessian was observed to slightly reduce the sensitivity of estimating thin CSF structures, though the concerns expressed about analysing downsampled smoothed Hessian results made at the start of this section must be considered. The potential accuracy of CSF detection was much higher than for BG, since the number of BG voxels in the ROI was very low in comparison and around half of these voxels were part of a poorly-contrasted vessel (marked with a yellow arrow in Fig 3.12). Therefore, some BG could only be detected by increasing the allowed false positive rate. Furthermore, the effects of any small mistakes in the expert labelling would be magnified compared to the more populous CSF.

Notable observations on the results in Fig 3.12 include high values of T^{CSF}_{new} being present at the edge of the lateral ventricle (marked with a pink arrow) despite its large size, and also a thin band of BG being flagged adjacent to this edge in T^{BG}_{new} (aqua arrow)[8]. Dark intensities (termed "ghosts") were indeed apparent here in Fig 3.12F, despite this being a GM region with no BG present. After consultation with Dr Farrall it was determined that no blood vessels were likely to be present around these ghosts, and conjecture as to the reason for the dark band's existence included motion artifacts or an artifact associated with local displacement of intensity by hyperintense CSF. Moving on to the errors present in the thresholded thin structure map B, a small patch of BG (orange arrow) has been wrongly flagged in a local intensity minimum caused by a sulcal interruption, and the topmost vessel is well contrasted in T^{BG}_{new} (brown arrow) but the more poorly-contrasted vessel below it (yellow arrow) is not. The latter vessel actually has a stronger value in T^{CSF}_{new}; this occurred due to this region lying on the boundary of the large lateral sulcus, causing it to contain partial volumes of CSF.

β_{new} was strongly negatively correlated with thin structure contrast in R^B_{new}, and overall the noisiness of the results in non-thin structure areas was much greater than for S_{new}. The thin structures highlighted were not just restricted to BG and CSF: in R^B_{new} calculated for the PD data, for example, the high-valued regions include a strip of WM running between the insula and putamen as well as thin WM regions in the insula (marked with orange arrows). However, the sensitivity to extreme intensity contrast in S_{new} meant that these areas did not manifest in T^{BG}_{new} (Fig 3.13H). Both vessels (pink

[8]Both of these regions were removed from B after applying the large contrasting structure removal processing outlined in Section 3.4.7.

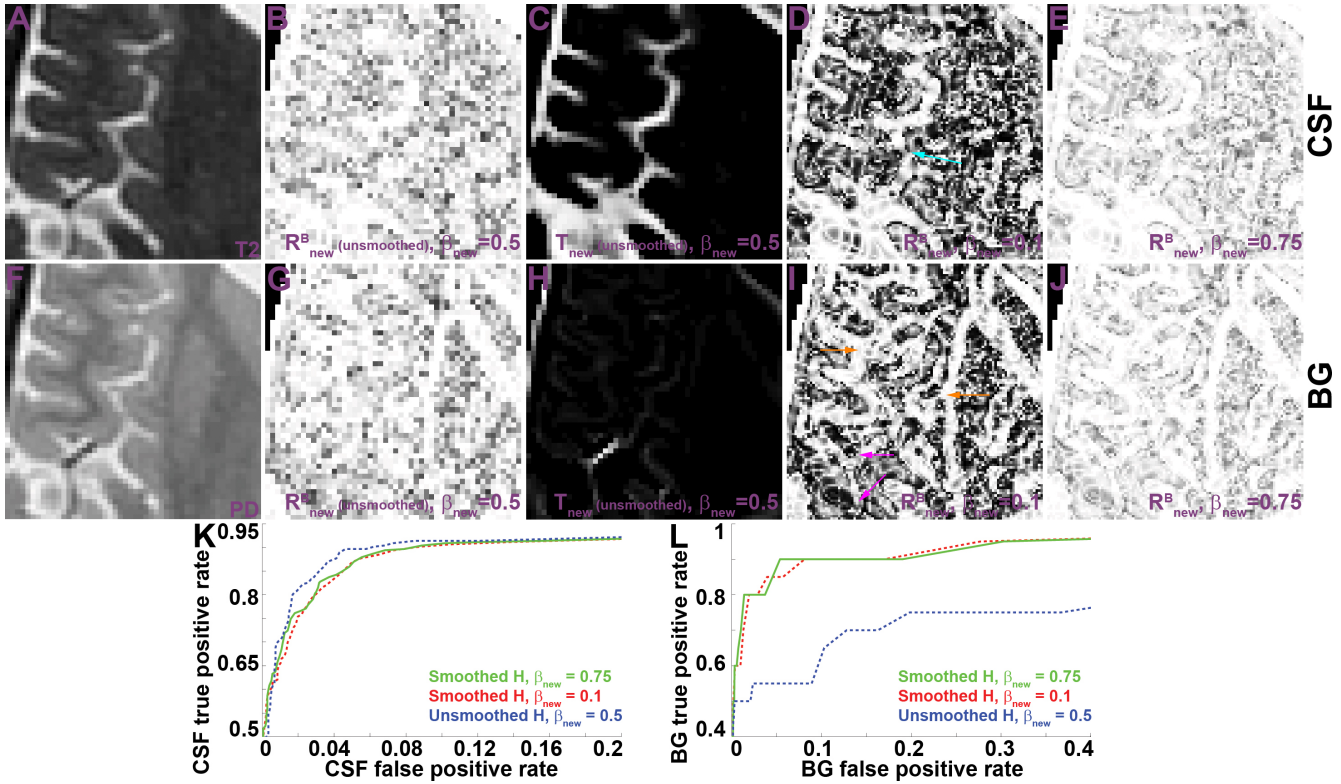

Figure 3.13: **Performance of new deviation of blob-shape measure R^B_{new}.** **A** & **F**) T_2 and PD data for ROI 5. **B** & **G**) Value of the R^B_{new}-containing term of T_{new} (Dfn 3.4.4) for **A** and **F** with $\beta_{new} = 0.5$. **C** & **H**) T_{new} for **A** and **F** with $\beta_{new} = 0.5$. **D** & **I**) Value of the R^B_{new}-containing term of T_{new} for **A** and **F** with $\beta_{new} = 0.1$. **E** & **J**) R^B_{new} for **A** and **F** with $\beta_{new} = 0.75$. **K** & **L**) ROC curves for **B**, **D**, **E** and **G**, **I**, **J** respectively. **unsmoothed**: no Hessian smoothing performed. Brighter grayscale values in panels B–E and G–J indicate greater values of the relevant measure. Coloured arrows are referenced in Section 3.5.1.

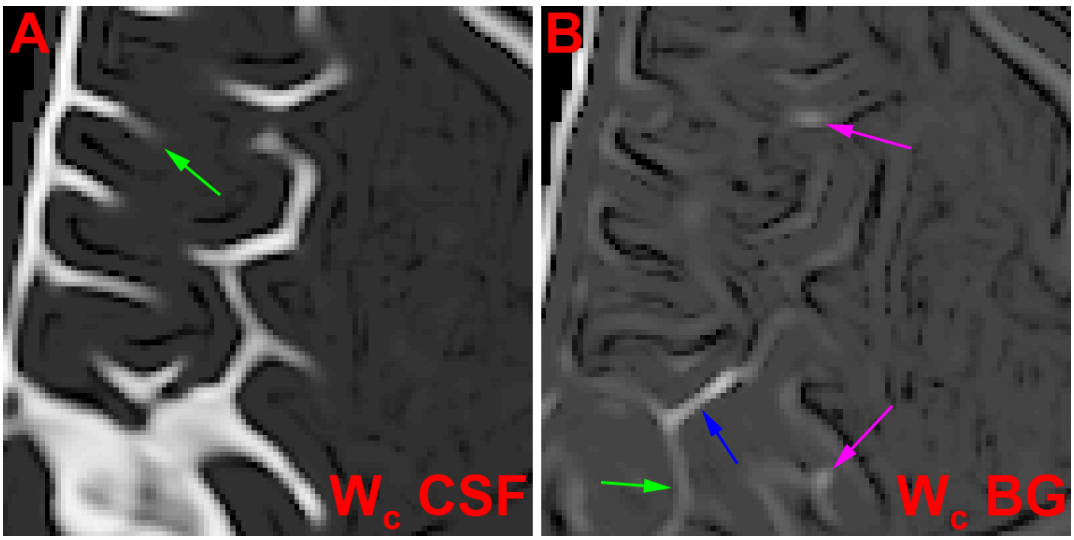

Figure 3.14: **Hessian certainty map W_c for Hessian smoothing**. **A**) W_c for the T_2 data in ROI 5. **B**) W_c for the PD data in ROI 5. Brighter grayscale values indicate greater values of W_c for the particular image weighting factor. Coloured arrows referenced in Section 3.5.2.

arrows) appear fairly bright in R^B_{new}, though the bottom one does not appear bright in T^{BG}_{new} due to its poor response in S_{new}. Hessian smoothing had a more dramatic effect on R^B_{new} than S_{new} as both the contrast of and uniformity along thin structures were much improved with this step, even when considering a higher value of β_{new} (Fig 3.13B, E, G and J). Using the smoothed Hessian induced a remarkable 30% absolute improvement in sensitivity at low false positive rates for BG (Fig 3.13L), due to the stark difference in sensitivity between R^B and R^B_{new} for the bottom vessel.

Considering the thin CSF structures, calculating R^B_{new} using the smoothed Hessian allows slightly fuller coverage (examined in more detail in Section 3.5.2) which is less sensitive to deviations in the local geometry due to poor resolution or noise (aqua arrow). The effects of analysing downsampled results previously appreciated when analysing Fig 3.12M should also be bourne in mind for Fig 3.13K; once again, a slight reduction in sensitivity at higher false positive rates was observed.

3.5.2 Further examination of the effects of Hessian smoothing

Correct formation of the data certainty map W_c (Dfn 3.4.7) determines the success of the Hessian smoothing process: it should possess high values only within thin structures (and highest at their medial axes) so their eigensystem properties can be propa-

gated toward their boundaries, and minimal values should be present on boundaries so the properties are not spread to other bordering tissues. These properties are clearly visible for most thin CSF structures in ROI 5 (Fig 3.14A), but in the PD-weighted data then local minima (marked with pink arrows) are flagged more often due to a lower maximum of S_{new} to produce fairly bright patches at some cortical boundaries with CSF. The topmost vessel (blue arrow) is bright as expected, but it does not possess a complete border of minimal values due to smooth changes in the local angular component of the Hessian (likely as a result of PVE). Similar problems are observed for the poorly-contrasted vessel below it and the tips of some sulci (green arrows) which do not appear to have clear and distinct edges with the cortex.

In Section 3.5.1 Hessian smoothing was found to improve the contrast and uniformity of R^B_{new}, together with the extent of high values of T_{new} along thin structures. The latter effect was found to occur at the penalty of a slightly increased false positive rate in the analysis at the downsampled level. The effects of this smoothing on the computed axial in-plane principal normal directions and the thresholded map coverage are contrasted in Fig 3.15; it should be noted that the smoothed Hessian results presented are effectively sub-voxels with co-ordinates located just inferior to those in panel A, and naturally there will be some under- or over-estimation of border locations due to the 3D nature of structure boundaries in the through-slice direction. In addition, since the large CSF region removal procedure is tuned for upscaled data then flagged voxels inside the lateral ventricle in the top right corner of Fig 3.15A should be disregarded. This procedure has also removed the region surrounding a widely-flaring sulcus (black arrow) in the smoothed Hessian results, suggesting that the morphological operations outlined in Proc 3.4.2 may need further tuning. Similar issues can also be seen in volumetric renderings of 13 slices of T^{CSF}_{new} surrounding ROI 5, presented in Fig 3.16, processed both with and without Hessian smoothing.

In concordance with the previous section's observations, the topmost vessel visible in Fig 3.15 is identified (marked with an orange arrow) with slightly improved uniformity of its principal normal directions in the smoothed Hessian result, though the vessel just below it (brown arrow) is not in either panel. In both results the dark skull (pink arrow) next to the edge of the parenchyma mask is correctly identified as BG; the principal normal directions can correctly run along the edge or in the through-slice direction since a tubular model is being enforced upon a plate-like object. Erroneous identification of BG only occurred at a dark group of cortical voxels between sections of a sulcus (blue arrow), and the isolation of these false positives suggests a

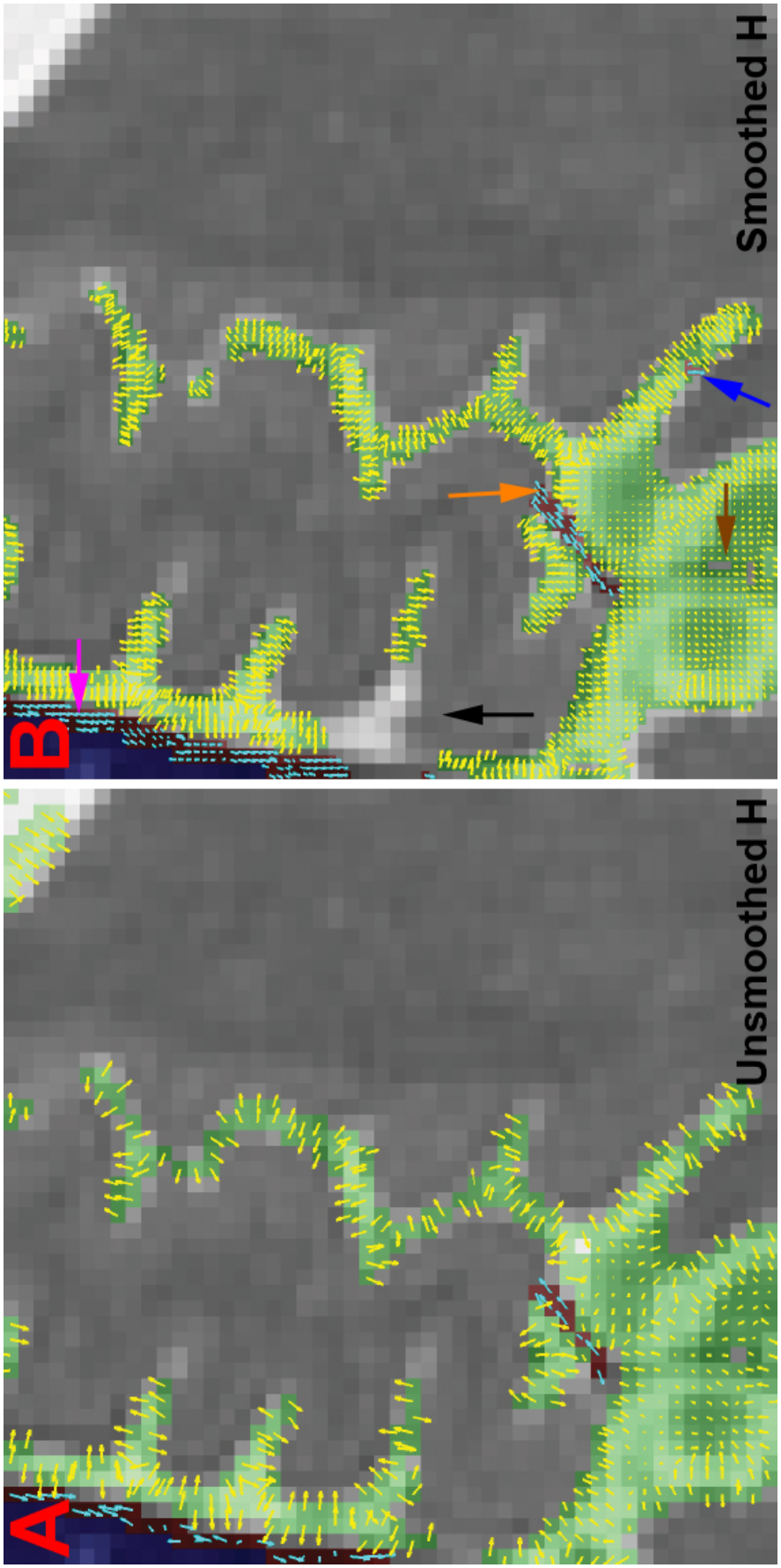

Figure 3.15: **Visualisation of the effects of Hessian smoothing**. Brain mask labelled in blue, tube-like structures labelled with red, and plate-like structures labelled with green for PD data for ROI 5 with (**B**) and without (**A**) Hessian smoothing. Yellow and cyan arrows denote the in-plane principal normal direction for plates and tubes, respectively. Coloured arrows referenced in Section 3.5.2.

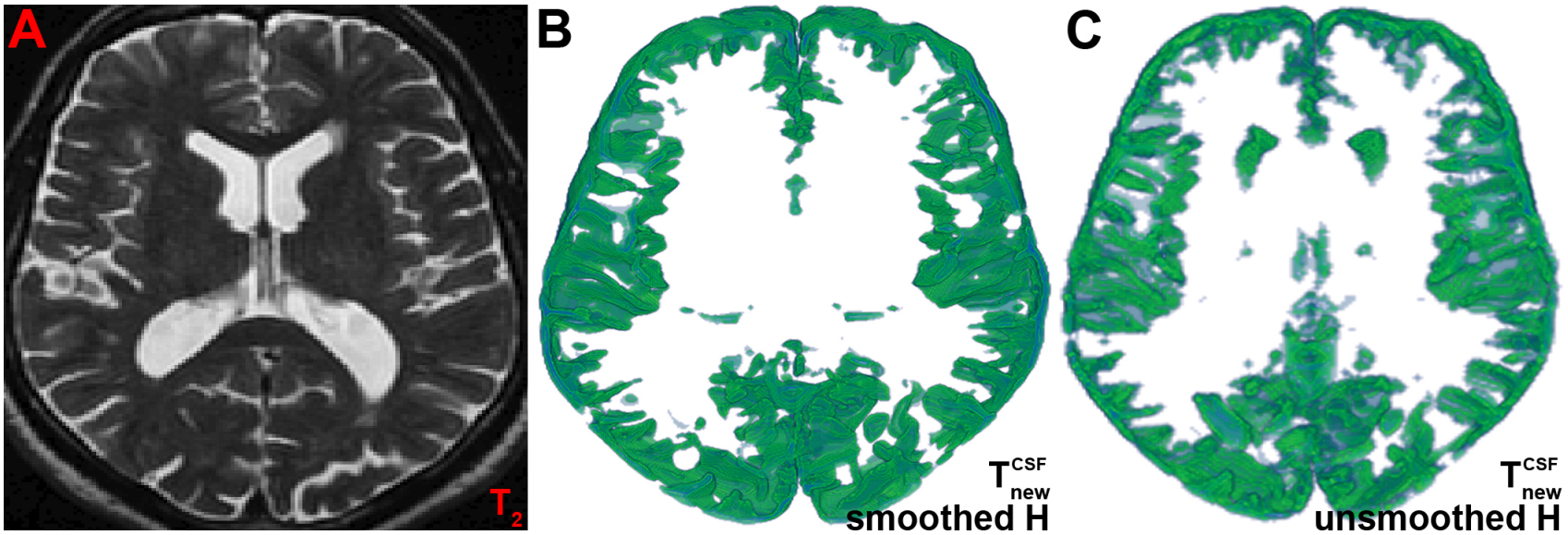

Figure 3.16: **3D rendering of Hessian smoothing effects**. Simian-rendered data is shown for 13 slices surrounding ROI 5 in panels B and C. **A**) T_2 axial slice on which ROI 5 is located. **B**) Rendering of T_{new}^{CSF} without Hessian smoothing. **C**) T_{new}^{CSF} with Hessian smoothing. Some remnants of the ventricles are present in the middle of **B** and **C**.

connectivity-based or structural model for vessels and sulci may be beneficial. Overall the sulci and CSF at the edge of the cortex appear to be more fully estimated in the smoothed Hessian result (after examination of several slices surrounding that pictured in Fig 3.15B) with more uniform principal normal directions.

Results from four more regions processed using Hessian smoothing are presented in Fig 3.17, which highlight more successes and failures. Regions of the ventricles highlighted with black arrows have not been removed by the large CSF region removal procedure due to their thinness of shape; if these regions are definitely not valid targets for future adaptive processing operations, then this results is motivation for further refinement of the method. In Figs 3.17A and C the extents of some poorly-resolved vessels (marked with aqua arrows) were not fully captured by the thresholding, although thicker and better-resolved vessels (blue arrow) have been more fully covered. Sulcal coverage is excellent in Fig 3.17A and small – likely pathological – blob-like regions of CSF in the caudate (brown arrow) are ignored due to the insensitivity of R^B_{new} to blobs.

Performance appears adequate in the cerebellum (Fig 3.17B) where there is very rapid change of tissue content between WM, GM and CSF in the extremely convoluted cortex. Two large regions of CSF have been successfully removed by the large CSF region removal procedure (green arrows), but some very poorly-contrasting sulci have not been flagged (purple arrow). A medial-occipital region pictured in Fig 3.17D has highlighted a long BG region attributed to the falx cerebri but only a small amount of the surrounding CSF, which was poorly contrasted and very much affected by PVE in the T_2-weighted data; furthermore, narrowing of the falx cerebri causes even more reduced contrast and in turn forms a break in the mapping (orange arrow). Finally, a dark intensity ghost (pink arrow) appears below the lateral ventricle, but it was not flagged as BG due to its elimination via the large CSF region removal procedure.

3.5.3 Quantification of errors in thin structure detection

The full summary statistics for the performance of the proposed methods on the expert-labelled LBC1936 ROIs are listed in Table 3.1[9]. ROIs 2, 4 and 6 did not contain any BG, and ROI 6 only contained a large CSF region (a ventricle), and so these regions were excluded from testing (marked N/A). Excluding the ROIs mentioned below which have substantial issues with the expert labelling or with the action of the large CSF

[9]In Figs 5.10-5.15, plots of the upscaled computed maps and the expert labelling at the original resolution for the first six ROIs can be inspected in panels I, L, M and N.

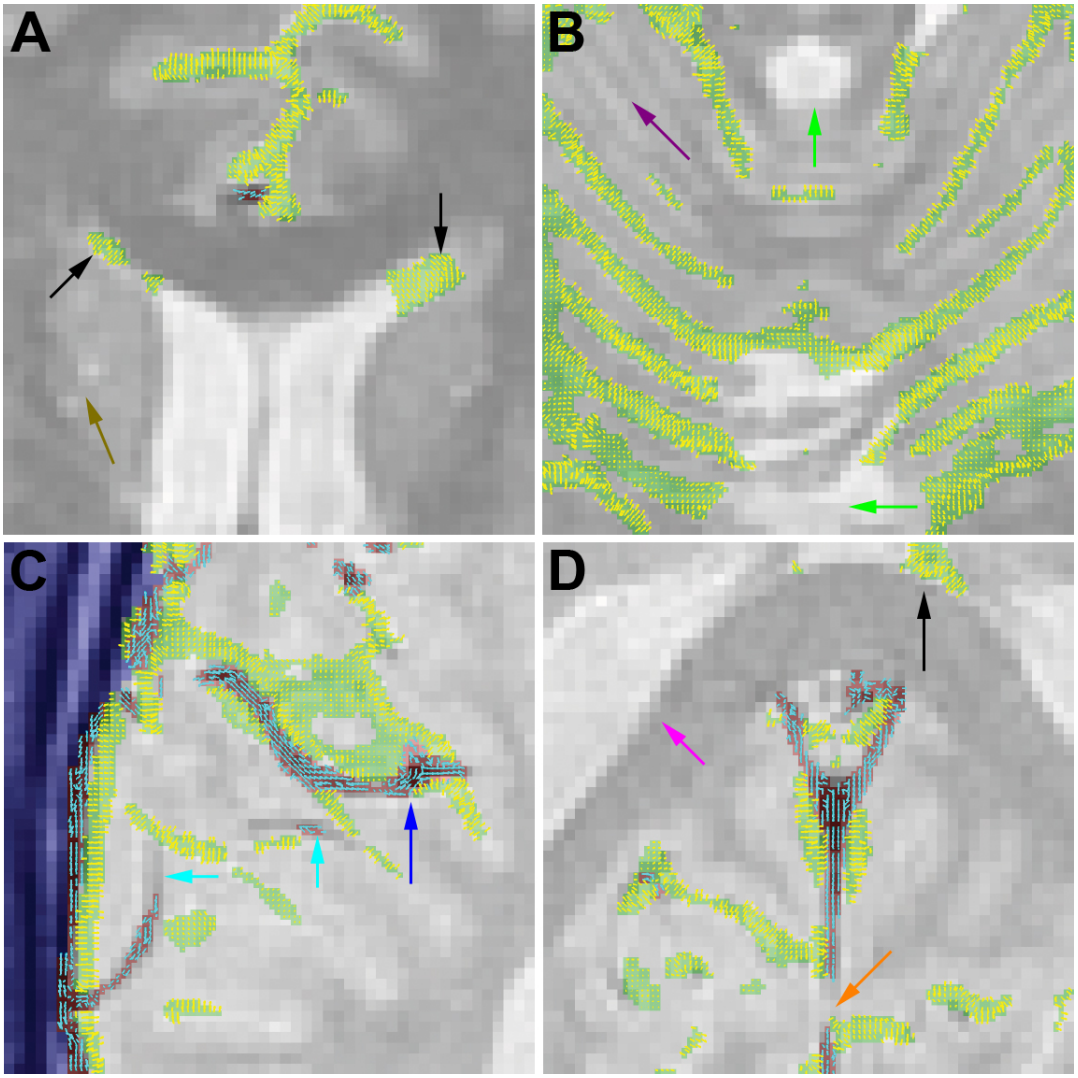

Figure 3.17: **Thresholded regions and principal normal directions for thin structure detection over several ROIs**. The parenchymal mask has been labelled in blue, tube-like structures in red, and plate-like structures in green over the PD data for each ROI. Yellow and cyan arrows denote the in-plane principal normal direction. **A**) ROI 1. **B**) ROI 4. **C** & **D**) Two ROIs from the young brain dataset. Coloured arrows are referenced in Section 3.5.2.

	CSF							*BG*						
ROI	**Sens.**	**Spec.**	**PPV**	**NPV**	κ	ρ	**Vol. error**	**Sens.**	**Spec.**	**PPV**	**NPV**	κ	ρ	**Vol. error**
1	56.90	99.54	93.40	95.28	0.68	0.82	-40	77.78	100	100	99.88	0.87	0.79	-9
2	86.34	89.86	84.31	91.25	0.76	0.70	-329	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3	32.72	99.00	95.04	71.59	0.37	0.58	-320	15.25	99.81	69.23	97.63	0.24	0.67	-58
4	59.81	88.5	81.84	71.76	0.49	0.48	-719	N/A	N/A	N/A	N/A	N/A	N/A	N/A
5	77.53	97.84	88.99	95.08	0.79	0.84	-62	25	100	100	99.31	0.40	0.62	-20
6	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7	19.51	100	100	93.37	0.31	0.66	-103	37.5	100	100	99.67	0.54	0.68	-8
8	8.77	99.95	83.33	97.44	0.15	0.19	-61	50	100	100	99.9	0.67	0.81	-4
9	78.17	98.43	93.10	94.35	0.81	0.88	-75	6.67	100	100	99.34	0.12	0.54	-1
10	42.57	100	100	78.03	0.50	0.73	-398	63.64	99.90	77.78	99.80	0.70	0.68	-18

Table 3.1: **Error rates for thin structure detection at each expert-labelled ROI**. **NPV** = negative predictive value. **Sens.** = sensitivity. **Spec.** = specificity. **PPV** = positive predictive value. Errors are listed to 2 decimal places. κ = κ-statistic score. ρ = correlation co-efficient between T and the ground truth. **Vol. error** = sum of binary differences from the ground truth.

region removal procedure, the results achieved their goals of reaching high PPV and specificity ($> 95\%$). In addition, the correlation between the continuous value of T_{new} with the partial (assumed 50%) and full (assumed 100%) content of the appropriate tissue type in the expert labelling was generally high, with the co-efficient ranging from 0.48 to 0.84. BG voxels only comprised a small fraction of the total voxels in each ROI and so naturally the NPV was always extremely high.

Concerns raised about reduced sensitivities and the κ-statistic scores when downsampling the smoothed Hessian results for quantitative comparisons, expressed at the start of Section 3.5.1, are again relevant to the following analyses and are very apparent for BG. Every volume error was negative, and attempting to remedy the putative thresholding issue by, for example, reducing the BG threshold π^{BG} to 0.03 had little impact on the specificity and PPV but increased the sensitivity in ROI 5 to 50%. The sensitivities and κ-statistic scores for CSF were generally better since sulci were generally larger than vessels and thus less affected by the effects of downsampling at their borders. CSF content was greater in the ROIs than BG, and accordingly the magnitude of their volume errors was larger and the NPV displayed a wider range of values. ROIs 7 and 8 chiefly contained ventricular components with some adjacent CSF structures that were mostly removed by the large CSF region removal procedure, explaining the low sensitivity.

Results from two ROIs bear particular mention for extreme discrepancies with the expert labelling. ROI 3 presented a very complex region of temporal cortex containing blood vessels, many large sulci, and pathological blobs of CSF in the GM. Many of the thresholded CSF sub-voxels were removed by the large CSF region removal procedure, and for some remaining large sulci then their core regions were extracted but not their edges. This last effect may have occurred due to considering a range of scales which did not encompass those relevant to these larger sulci; however, the impact of this lack of sensitivity is lessened since it is expected that these larger sulci will be easier to process than their thinner and more poorly-resolved counterparts in future chapters. More blood vessels appeared in this ROI than the expert labelling suggested when slices above and below the pictured slice were investigated – explaining some of the false positives – though the vessels flagged by the expert have poor contrast and thus contributed to the low sensitivity. ROI 10 is notable for its low CSF sensitivity and non-perfect PPV for BG. These statistics can be explained by the presence of many low-contrast sulci in this occipital region and some thresholded voxels being removed by the large CSF region removal procedure, as well as by the absence of BG at the top

of the falx cerebri in the expert labelling.

3.5.4 Effects of shaped and oriented filtering

Fig 3.18 presents a simple application of the shaped and oriented anisotropic filters (generated using the methods detailed in Section 3.4.4) in smoothing ROI 5[10]. Isotropic smoothing reduces the intensity extrema of the sulci and vessels (black arrows in Fig 3.18C and D), but the adaptive smoothing result is able to prevent some of the intensity loss for bright sulci (red regions in Fig 3.18E and F) and also prevent some intensity gain for the dark topmost vessel (blue in Fig 3.18E and F). The latter effect is not repeated for the vessel below it since it was not flagged by the BG detection procedure and it was instead flagged as CSF on the sub-voxel level slice, hence explaining its increase in intensity after smoothing with a plate-shaped filter across it.

3.5.5 Processing efficiency

An important requirement of the method is that it can process a full volume using a modern desktop PC within a few hours. One external code source was used: the calculation of the eigensystem of 3×3 Hessian matrices was performed using the LAPACK[11] routine `SSYEV`. This code requires compilation for the target platform and is suitable for multi-processor environments, but no precise quantification of its computational efficiency for the given problem size could be found in the literature or elsewhere.

Over half of overall processing time was spent in smoothing the Hessian (described in Section 3.4.5) for which the physical dimensions of the Gaussian filters used – generating the local average of the Hessian H_{nhood}, the scale-normalised gradients (Eqn 3.6), as well as forming the applicability matrix W_a – had a major impact. However, if most arithmetic operations can be considered to cost a multiple of the input data size, then the increase in requirements for storage space and processing complexity for the rest of the operations performed (excluding `SSYEV`) can also be considered linear in this factor. One exception could be upscaling the data, performed via cubic interpolation using MATLAB's `interp3` function which took around a quarter of overall processing time to complete but was only performed only once at the start of processing. The

[10] The thin structure map and computed principal normal directions for this ROI can be referenced in Fig 3.15.

[11] Available from `http://www.netlib.org/lapack/`.

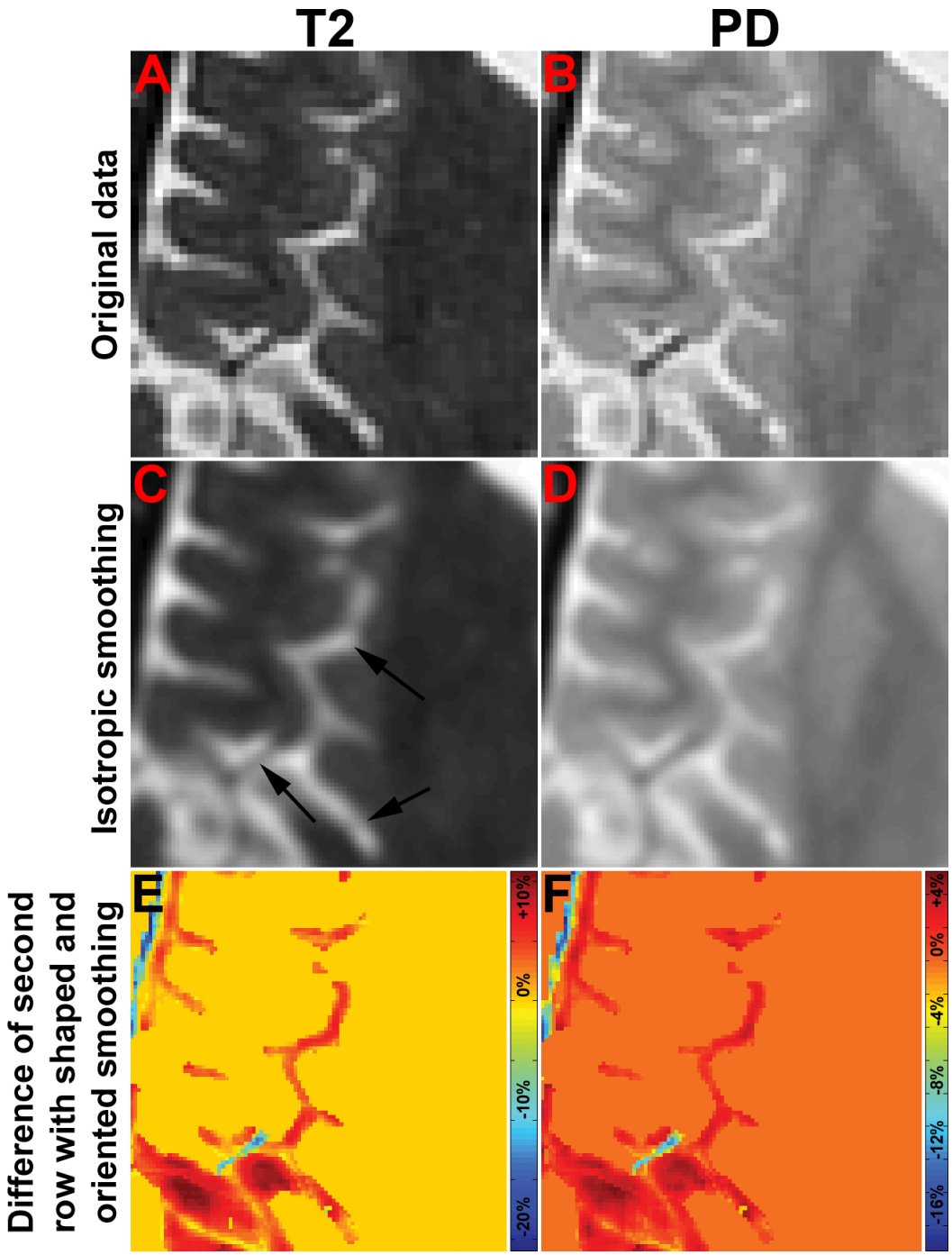

Figure 3.18: **Effects of shaped and oriented smoothing**. **A** & **B**) T_2 and PD data for ROI 5. **C** & **D**) **A** and **B** after isotropic smoothing with filter size $5 \times 5 \times 5$ and $\sigma = 1$. **E** & **F**) Difference of shaped and oriented smoothing (with $large$ to $small$ ratio 2:1) from isotropic smoothing applied to **A** and **B**. Percentage differences shown are relative to the maximum intensity of **A** and **B**, respectively. Black arrows referenced in Section 3.5.4.

upscaling ratio L determines the input data size and also the filter sizes used in Section 3.5 were a function of L, so this factor should be minimised to increase processing efficiency (likely at the cost of accuracy).

Over three trials of processing the entire volume containing ROI 5 as described in the introduction of Section 3.5, the time taken was 4801.4 ± 27.3 seconds (mean $\pm$ standard deviation). Of course by investigating more scales, using larger filters, resizing the volume by greater L and processing higher-resolution datasets then this time will be increased and memory requirements will rise. The processing time can increase very rapidly when very large matrices are being created which exceed the amount of RAM available; for example, when the resizing ratio $L = [3,3,3]$ is used then significant amounts of time will be spent paging data from the hard drive on a modern desktop PC with 4GB RAM.

3.6 Discussion

Referring to the objectives stated in Sections 3.3 and 3.4, the proposed methods present a viable approach for detecting thin contrasting structures in dual-echo T_2-PD structural brain MRI acquired in clinical settings, chiefly comprising cortical sulci and blood vessels. The methods exploit the individual exceptional tissue contrasts of these image weighting factors but do not require additional scans to be acquired, nor otherwise increase the scan time. While blob-like structures – such as oedema present in GM observed in ROI 1 (Fig 5.10) – induce a poor response, otherwise there is no discrimination between the local tube- and plate-shaped intensity profiles that thin structures containing BG or CSF can take, which is also advantageous when considering data with a high slice thickness. Once the parameter training is complete then the method is automated, it has no reliance on template matching to brain atlases which may not be representative of these unique components of the anatomy, and it does not require seed point placement nor other manual intervention.

High PPV for thin structure detection was attained with high thresholds of T_{new} (Dfn 3.4.4) to provide highly specific priors for the future resolution enhancement (Chapter 4) and segmentation (Chapter 5) stages of processing. However, this strategy contributed to poor sensitivity in the results listed in Table 3.1 for ROIs in the LBC1936 dataset, even taking into account reservations regarding performing analysis on downsampled results (expressed in the introduction to Section 3.5). Furthermore, the statistics produced were dependent upon the accuracy of the expert-labelled ground

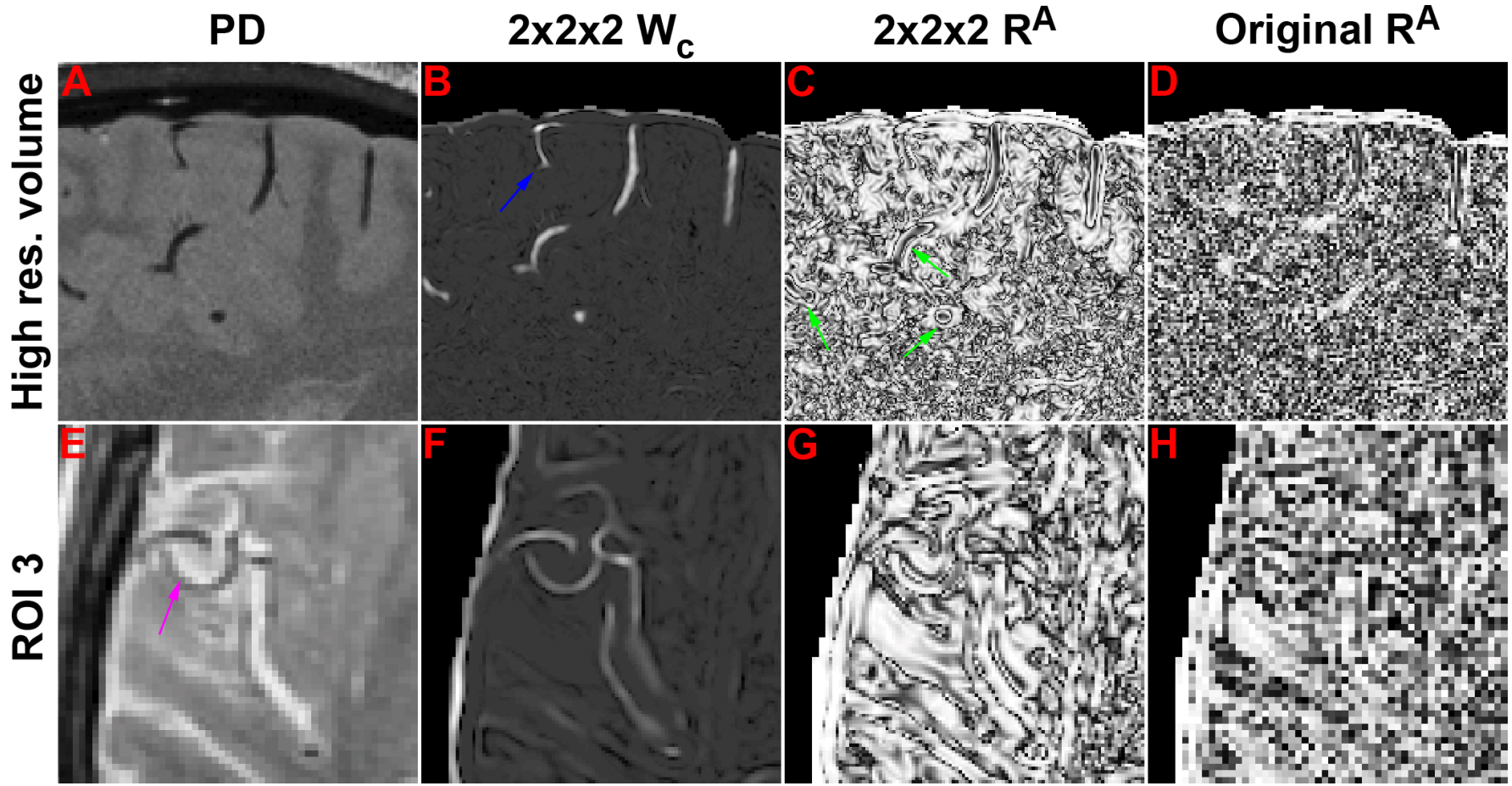

Figure 3.19: **Improvement in the quality of R^A and W_c using high resolution data**. **A**) High-resolution PD data. **B** & **F**) Hessian certainty W_c (Dfn 3.4.7) of **A** when full thin structure detection with upscaling ratio $2 \times 2 \times 2$ is applied. **C** & **G**) R^A of **A** when full thin structure detection with upscaling ratio $2 \times 2 \times 2$ is applied. **D** & **H**) R^A of **A** when thin structure detection with no upscaling is applied. **E**) PD data from ROI 3. **High res. volume**: high resolution volume. Coloured arrows referenced in Section 3.6.

truth, and in ROIs 2 and 10 it was noted that large BG regions in the falx cerebri were not labelled correctly. Most detection errors were found on thin structure boundaries where the labelling is expected to be most subjective, and thus to eliminate rater bias and mistakes then ideally the single-rater single-visit labelling approach should be replaced by a multiple-rater repeated-labelling consensus (proposed in Section 6.3).

Comparison of results displayed at different resolutions is problematic, though one solution – that does not involve sensitivity to interpolation nor require expert labelling of the individual sub-voxels – would be to use contour-based structural models fit to both results. This kind of comparison has been performed for larger brain structure [Yushkevich et al., 2006] using the Dice similarity co-efficient [Zou et al., 2004]. Application of this kind of modelling (covered in more detail in Section 6.4.1) to thin structures could also foster notions of continuity in both vessel and sulci networks, as well as local uniformity of thickness along their lengths (or at least slow changes). Explicit surface models may also improve smoothing of the Hessian to the structures' edges and also prevent isolated local intensity extrema – such as patches of noise (i.e. marked with an orange arrow in Fig 3.12H) – from being identified as thin structures, or as breaks in them.

The factor that seemed most to blame for the poor sensitivity was the high 2mm slice thickness of the datasets studied, which was twice the in-plane voxel dimensions. This conclusion was made from observations that both thicker blood vessels proceeding directly axially as well as vessels oriented in the through-slice direction (for example, pictured near the middle of Fig 3.17A and C) were detected well, but the contrast of thinner vessels proceeding axially (such as those in Fig 3.19E, pink arrow) and especially those with both axial and through-plane orientation components was poor. These vessels occupied only a fraction of the slice thickness and thus were affected heavily by PVE. Reducing the slice thickness – for example, by examining a higher resolution acquisition in Fig 3.19A and B – appeared to remedy the issue. Further theoretical support for this argument comes from the analysis of the Gaussian second derivative filters (pictured in Fig 3.6) used to calculate the Hessian: high slice thickness means "through-plane" filter element weights will be minor compared to those in the central "axial slice" of the filter, and the intensity extrema expected at the central element of the filter will be reduced in magnitude by PVE. This effect is less noticeable for sulci since they only possess one thin dimension, and because they are generally thicker than vessels.

To remedy this issue then the thresholds π^{BG} or π^{CSF} can be lowered – which will

result in more false positives (from thin regions of WM in the case of vessels) – or the acquisition protocol for the MRI data could be changed. Making the WM class in PD brighter in intensity would somewhat alleviate this confounding factor for detecting vessels, though a more satisfying solution would involve re-designing the scan sequence to either increase BG contrast or to reduce the slice thickness to at most the thickness of a typical vessel. As a consequence, vessels would appear much better contrasted – even for very thin vessels pictured in Fig 3.19B (blue arrow) – and more accurate detection would be possible. Segmentation methods which optimise their labelling in the context of neighbours (investigated in Chapter 5) could also be employed to help discriminate vessels since CSF only usually borders WM around the ventricles or in pathological regions, meaning dark intensities bordering CSF elsewhere are likely to be BG.

Hessian smoothing (proposed in Section 3.4.5) helped to maintain similar values of T_{new} along thin structures and the Gaussian filters used ensured the maximum principle was observed, with the greatest values still present at the structures' medial axes. The conditions found at the medial axes were propagated towards the structure edges, as observed in slightly fuller coverage of sulci and vessels in Fig 3.15, and improved uniformity of the principal normal directions for thin structures could also be observed in this figure. Other notable effects of Hessian smoothing included increased contrast of the plate-tube shape discriminant measure R^A (between structures marked with green arrows in Fig 3.19C and D) as well as the increased contrast of R^B_{new} (Fig 3.13). Both these results suggest that the smoothing effects affect even the smallest components of the eigensystem used in the numerator of both measures (Eqns 3.7 and 3.16).

This chapter described two procedures aiming to increase the specificity of thin structure detection, which had mixed success. Firstly, the large contrasting structure removal process (proposed in Section 3.4.7) removed the majority of the ventricular content and prevented dark "ghost" bands of intensity next to their edges (encountered in Section 3.5.1) from being flagged. However, the thin horns of the lateral ventricle were not removed and the bases of widely-flaring sulci near the cortex were incorrectly removed. These problems can be resolved by restricting the regions removed by a minimum connected component size, and also by using a region-growing method to include the ventricular horns within the large region mask. Secondly, a procedure designed to select relevant scale-space levels to investigate (Proc 3.4.1) used a simple, practical and fairly fast 2D method to sample and measure thin structure thicknesses – and it was easily implemented by modifying the manual labelling assistance software

(described in Section 2.4). This method demands validation in other datasets, as well as extension to 3D processing (necessitating more advanced visualisation and labelling tools) and spline-based structure length measurement to more accurately quantify their thicknesses. This last modification will require re-evaluation of the relationship computed between optimum scale-space levels and measured thin structure weaknesses (proposed in Section 3.4.6).

In the literature the vessel enhancing filtering approach of Frangi *et al*. has only rarely been used on structural MRI data and has not been applied to detecting cortical sulci, most likely due to the relatively poor contrast of blood vessels and sulci to other tissues when comparing structural MRI to other imaging techniques such as MRA. The filter has been applied to highlight vessels in PD data [Descoteaux et al., 2008][Descoteaux et al., 2004] while also using flux-maximising geometric flow [Vasilevsky and Siddiqi, 2002] to drive T (Dfn 3.3.3) toward vessel boundaries from extracted medial axes. While in these works similar propagation strategies to the methods proposed in Section 3.4.5 were performed, a difference emerges in the formulation of a stopping condition for smoothing which is based on the local intensity gradient rather than the local certainty of thin structure content in the Hessian data (Dfn 3.4.7). The utility of these conditions will degrade with PVE at vessel boundaries; however, by working at the sub-voxel level and also considering a compound measure examining both local angular similarity of the Hessian as well as the magnitude of maximal local curvature then the localisation of these boundaries could be better using the proposed methods. Nonetheless, the agreement of the method presented in [Descoteaux et al., 2008] with manual labelling was impressive – reported κ-statistic scores ranged from 0.7 to 0.8 on a publically-available corpus of 20 normal brains [Aubert-Broche et al., 2006a]. While these statistics appear to outperform ones listed for BG in Table 3.1, these volumes which were processed were acquired with isotropic dimensions (putting the testing of the proposed methods on thickly-sliced data with much more PVE at a disadvantage). Future validation work should examine the performance of vessel and sulci detection on this standard dataset. Some of the results presented in [Descoteaux et al., 2008] compared a minimum intensity projection of volume intensities (since the vessels had a dark intensity) with a maximum intensity projection of the vessel map (where highlighted vessels were bright), and this visualisation method should be considered when presenting results from volumetric data, in addition to tensor visulation methods for plotting the second-derivative ellipsoid (such as that presented in [Criscione et al., 2000]).

Fewer parallels can be drawn between Hessian smoothing and vessel enhancing diffusion [Manniesing et al., 2006] since this method is iterative and requires termination criteria, and also because the intensities are smoothed rather than the Hessian so the smallest components of its eigensystem will not be directly affected. Furthermore, this approach was designed for MRA data which has superb contrast for blood, and so there was little need for thin structure certainty maps to constrain the smoothing. However, anisotropic diffusion bias along the principal eigenvectors has a similar goal to the adaptive applicability matrix W_a in encouraging smoothing along thin structures, and it also goes one step further in implementing dependency between smoothing anisotropy and the magnitude of the principal eigenvectors λ. A functional relationship between λ and the shapes in W_a as well as in the filters produced in Section 3.4.4 could help to promote better responsiveness to thin structures with rapidly-changing paths along their medial axes, as well as further reduce the effects of noise for less complicated ones; this approach has also been advocated in more general adaptive anisotropic smoothing methods (i.e. [Svensson et al., 2006]).

Phase congruency [Kovesi, 1999] was investigated in Section 3.3 as an alternative means for thin structure detection, but it was not considered further following observations of many false positives and branching between highlighted regions. Its most attractive qualities were its robustness to intensity contrast changes and sensitivity to orientation coherence, which suggests it may detect poorly-contrasted vessels and sulci better than S_{new} (Dfn 3.4.2). Indeed, as the magnitude of S_{new} was determined by intensity contrast and was very similar to T_{new} (as observed in Section 3.5.1) then it is unsurprising that there were a good correlations of T_{new} with the expert-labelled tissue content in Table 3.1. Further development of T_{new} could focus on the replacement of the S_{new} component with a 3D phase congruency-based measure, and using the deviation from blob-shape measure R^B_{new} (Dfn 3.4.3) to temper its tendencies to overestimate and branch between content. Furthermore, similarities in function between R^B_{new} and the absolute magnitude of the tensor mode developed in [Criscione et al., 2000] (examined in Section 3.3) – where blobs produced a zero value, and plates and tubes led to +1 and -1 values respectively – should warrant examination of the sensitivity of the tensor mode for this purpose.

Explicit surface-based modelling and atlas-based segmentation of cortical sulci have been popular topics of research, manifesting in FreeSurfer [Dale et al., 1999][Fischl et al., 1999] (examined in Section 3.3). The approach of using identified sulci to help infer the location of cortical GM is examined in more detail in the Conclusions

chapter (Section 6.4.1), but FreeSurfer makes use of a reverse methodology. The addition of GM-based boundary constraints – employing models of cortical thickness – to the sulcal identification procedure could help to further regularise the result in the presence of noisy intensities. However, since WM/GM contrast is often worse in dual-echo T_2-PD data compared to T_1-weighted data, then such explicit cortical modelling may need to be coupled with edge enhancement and volume smoothing methods (such as those presented in Chapter 4) alongside a dual-channel segmentation algorithm with more advanced neighbourhood optimisation (as proposed in Chapter 5).

The processing time for a sub-volume recorded in Section 3.5.5 was fairly long, but it was noted that computation of the eigensystem from the smoothed Hessian was particularly costly. All the filtering operations could be considerably reduced in cost by employing steerable filter-based systems [Freeman and Adelson, 1991] to construct the output from basis filters and avoid the discretisation of the orientation space. Furthermore, re-coding the non-LAPACK[12] operations involved in Hessian smoothing as MEX files may also be fruitful in reducing processing time. There is also considerable potential for parallelised computation to make an impact on performance. Identical processing occurs at each scale-space level and most matrix operations (with the notable exception of finding the local angular similarity c_a in Dfn 3.4.7) can be split into and processed as sub-matrices. The computation of many intermediate matrices can also be performed simultaneously.

As MRI scanning technologies improve in the future then either reducing noise, acquiring volumes with voxels possessing isotropic dimensions or minimising the scanning time will likely be preferred in clinical practice over increasing the axial-plane resolution. Blood vessels and cortical sulci thus may remain somewhat poorly-resolved for some time to come and their specific detection – allowing adaptive processing – will still be worthwhile. Many thin structures such as the falx cerebri remain poorly-resolved even at high 512×512 axial-plane resolutions, and at this level and smaller vessels and sulci will become visible, and so there is still scope for thin structure detection to remain a useful processing step even as resolution improves. Furthermore, increased resolution and reduced noise should allow better definition of the Hessian, as well as improved discrimination of blood vessels. Consequently, the plate-tube discriminant R^A will become more viable for dividing the membership of B^{BG} into tubular blood vessels and other plate-like background structures (such as the meninges) in a post-processing thresholding step. An initial investigation of this discriminant (Fig

[12] Available at `http://www.netlib.org/lapack/`.

3.19C) on the high resolution brain dataset was able to correctly assign minimal values of R^A to vessels (marked with green arrows) and maximal values to the skull (at the top border of the parenchyma), in a manner superior to that achieved with processing without Hessian smoothing (Fig 3.19D) and at lower resolutions (Fig 3.19G and H).

3.7 Summary

In this chapter a new method has been introduced for the detection of the presence, orientation and shape of thin contrasting structures in dual-echo T_2-PD structural MRI, with a high PPV. Following an introduction describing Gaussian scale spaces as well as the pitfalls of non-adaptive smoothing on small-scale structures (Sections 3.1 and 3.2), a review of existing techniques for the detection of blood vessels and cortical sulci has been presented in Section 3.3. Existing differential geometry techniques for blood vessel detection in MRA volumes proposed in [Frangi et al., 1998] have then been adapted to also detect cortical sulci in Section 3.4. Furthermore, tensor field regularisation methods inspired by Westin and Knutsson [Westin and Knutsson, 2003] have adaptively smoothed the Hessian tensor generating the thin structure maps in order to propagate conditions present at the medial axes of thin structures towards their boundaries, so that fuller and less noisy maps can be generated with higher thresholds.

The testing of the methods detailed in this chapter has mainly been accomplished through 2D and 3D visualisations of the thin structures and their defining shape vectors on a range of real MRI data. In the Discussion (Section 3.6) suggestions have been made to improve the visualisation of results still further and extend the quantitative analysis to the vectors as well. Better sensitivity of the highly-specific maps of the thin structures could be induced by incorporating components based upon phase congruency [Kovesi, 1999] or explicit surface-based modelling, and the description of local curvature could be enhanced by using measures such as the tensor mode [Criscione et al., 2000].

In Chapters 4 and 5 the resulting outputs from the methods described in this chapter are employed to create filters to help preserve the thin structures' anisotropy in subsequent smoothing and also in methods relying on neighbourhood context, to help prevent loss of volume and blurring. Furthermore, the thin structure maps provide valuable priors of tissue content that can be directly applied during segmentation and can also guide thin structure intensity restoration during resolution enhancement.

Chapter 4

Volume Resolution Enhancement

An original segmentation method for brain MRI volumes proposed in [van Leemput et al., 2003] aimed to simplify the soft labelling and class parameter re-estimation steps by working at an artificially-induced high resolution (a level of representation with sub-voxel accuracy). From a hidden Markov modelling perspective, observable MRI volume data where PVE could occur were modelled as being downsampled from this unobservable high resolution volume which was assumed to be free from PVE, and thus these sub-voxels should only be assigned hard labels. However, this restriction is too strong for general application to brain volumes since many structures may be poorly characterised using this model at this level: these include deep GM nuclei (where WM and GM appear to mix without interface at typical clinical MRI resolutions) together with many blood vessels, thin cortical sulci and tightly-compressed regions of WM between cortical gyri.

In order to resolve this modelling issue then a modified segmentation method – presented in Chapter 5 – will allow soft labels to be assigned at the sub-voxel level, but the problem of how to estimate the sub-voxel intensities remains. Coupling the resolution enhancement and segmentation steps in a unified process (as advocated by Van Leemput *et al.*[1]) makes the accuracy of the resolution enhancement step sensitive to several factors including the class parameter initialisation process as well as bias field effects (which effectively makes class intensity spatially-dependent). Not only does this lead to dependence upon the suitability of these models, but it also lengthens each iteration by performing volume resolution enhancement each time with respect to the current class parameters. Conversely, splitting the processes and calculating the

[1] A comparison of the new full segmentation method's results to that of [van Leemput et al., 2003] can be found in Section 5.4.8.

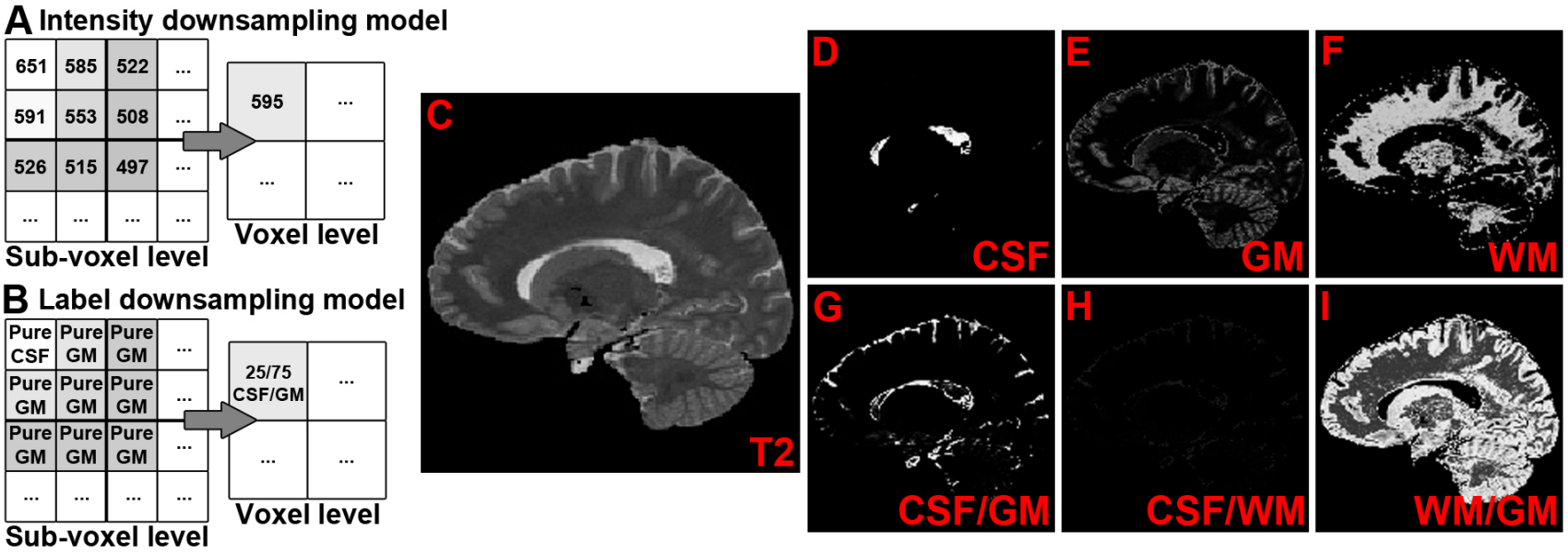

Figure 4.1: **Downsampling segmentation approach of Van Leemput** *et al.* [van Leemput et al., 2003]. The T_2 slice data and code were supplied by the author (available at `http://people.csail.mit.edu/koen/PVEM.html`) and the default parameter settings supplied were used. **A** & **B**) Intensity and label downsampling models for resizing ratio $2 \times 2 \times 2$. Thick boundaries at the sub-voxel level indicate voxel-level boundaries. **C**) Sagittal T_2 slice from a real brain volume. **D**-**I**) Particular label probabilities produced by segmentation (brighter intensities denote probabilities closer to 1).

sub-voxel intensities in a single pre-processing step before segmentation may alleviate these issues, but the result may lack reconstructive properties[2]. In this chapter the second option has been chosen due to its practicality, and constraints on the changes in sub-voxel intensities are enforced by analysis of neighbouring intensities as well as by exploiting the high PPV of the thin structure maps produced in Chapter 3.

The term *resolution enhancement* is used to describe the artificial upscaling of volumes rather than *super resolution* in order to distinguish the approach from well-known methods which rely upon multiple acquisitions of the same scene, sometimes with slight displacement of the target object (i.e. [Baker and Kanade, 2002] and [Nguyen and Milanfar, 2000]). Compared to traditional interpolation processes which assume a functional relationship between the required sub-voxel sampling points and their closest voxels, the proposed methods aim to exploit the different contrasts and noise characteristics of the dual-echo T_2-PD data to more accurately smooth homogeneous regions and sharpen the edges of tissue regions. Maximising the number of pure tissue sub-voxels produced in this fashion (thus reducing the proportion of the voxels affected by PVE) should simplify their later classification, and using flow and diffusion speed constraints helped to ensure that no extra features were created at the high resolution level that were not present at the original resolution.

Since the production of the thin structure maps (described in Chapter 3) was performed at an increased resolution, the dual-echo T_2-PD volumes' intensities were also upscaled to match. Localisation of the cortical sulci and blood vessels – which are generally poorly-resolved – and computation of their principal normal (Dfn 3.4.1) vector fields then allowed restoration of their contrasting (though still PVE-affected) interior pure tissue sub-voxels. These structures would be poorly handled by the original downsampling segmentation method of Van Leemput *et al.*: it would be extremely difficult to simultaneously reduce noise (low frequency intensity variation) whilst preserving these high frequency structures in the volume, where both occur on similar spatial scales.

This chapter is organised as follows. Previous approaches to structure-preserving smoothing and PVE reduction methods are examined first, with a detailed description of the adaptive anisotropic diffusion methods that form the starting point for the work in this chapter. The modifications to allow thin structure preservation and restoration

[2]These properties can be realised by the linear combination of class intensity means and their respective fraction contents at each volume location, using the mixel model of intensity formation [Choi et al., 1991], having also applied structural models or methods that ensure local label homogeneity during segmentation to provide robustness to noise, bias or other imaging artifacts.

as well as dual-channel resolution enhancement are then formally described in Section 4.2. Demonstrations of both the effects of these changes and also a validatory study of the proposed methods' application to volumes acquired at different resolutions follow, along with their discussion (Section 4.3). Throughout this chapter dual-channel data variables, which contain a 2-element vector at each point, are given an overline formatting style (i.e. $\overline{X}$) and iteration-dependent indices are written in superscript (i.e. X^t).

4.1 Survey of structure-preserving smoothing and PVE reduction approaches

To enhance the resolution and quality of structural MRI brain volumes, edge-preserving smoothing (or voxel intensity correction) must be combined with PVE reduction. The latter step necessarily involves the upscaling of the original volume to a resolution-enhanced version to enable the formation of sub-voxels at tissue boundaries with intensities similar to those of the neighbouring regions. The entire operation can be described in terms of non-linear filtering[3]: the relationship between the original and enhanced resolution versions can be defined using a degradation function, comprising an undersampling function and a blurring function. One component of the blurring function is known as the point-spread function (PSF), but this is not considered to be a major factor in MRI acquisitions (as claimed in [Salvado et al., 2006]) and hence deconvolution methods for extracting the PSF will not be investigated.

The other components of the blurring function – edge enhancement and homogeneous region smoothing – can be achieved using locally-adaptive processing. Key to these operations is the identification of edge and noise regions, for which wavelet-based approaches are a popular in the literature. Perhaps their most well-used properties are that the wavelet coefficient magnitude of noise is highly negatively correlated with scale due to the performance of the scaling function in the wavelet transform [Mallat and Hwang, 1992] so it is usually concentrated in the finest scales of wavelet-based scale spaces. The co-efficients for true object edges tend to be preserved over several scales, and in [Carey et al., 1999] discrepancies in the Hölder regularity across wavelet sub-bands were investigated for the purposes of edge detection during resolution enhancement. However, at the resolutions typically used in structural MRI by

[3]The reader is referred to review [Pollak, 2002] for broader coverage of the field of adaptive image filtering and restoration, and another review [Powell, 1996] for an overview of interpolation methods.

clinicians then many thin anatomical structures also exist at scales similar to those occupied by noise.

Variational approaches to image restoration use Bayes' Rule to maximise the probability of the restored data, given the original data and a functional model for the noise (such as additive zero-mean Gaussian). The resulting energy minimisation problem uses a fidelity term over certain features, which forces the result to remain close to the original in these respects, as well as a regularization term such as the popular Total Variation [Rudin et al., 1992] and Mumford-Shah [Mumford and Shah, 1989] functionals which prefer edge length preservation and intra-region smoothing. To avoid designing a new regularisation term for variational methods or having to re-work multi-channel wavelet reconstruction to deal with sub-voxel variation in dual-channel structure boundary locations and to discriminate between noise and thin structures, then diffusion-based methods for structure-preserving smoothing and PVE reduction have been selected. Their advantages in implementing PVE reduction, homogeneous region smoothing and thin structure restoration are discussed in the following sections.

4.1.1 Diffusion-based methods

In the context of processing brain MRI volumes, diffusion is an evolution of the intensities in a volume V, where intensity is not created nor destroyed but only transported between voxels according to differential equations driving this flux. Assuming that noise can be characterised by a zero-mean Gaussian distribution and that the mixel model [Choi et al., 1991] holds for MRI data, then diffusion appears ideal for PVE reduction inside voxels by redistributing sub-voxel intensities to match those of neighbouring regions. When performing diffusion, Fick's Law describes the motion of intensity over time t: $flux$ compensates for the intensity gradient $\bigtriangledown V$ as well as D, which is a positive-definite symmetric matrix specifying the diffusivity bias. The evolution of V can be described by the continuity equation shown in Eqn 4.1 which includes Fick's Law, where $\bigtriangledown \cdot$ is the divergence operator. Neumann boundary conditions create a sealed environment[4] so diffusion is halted across the boundaries of V.

[4] Also known as an adiabatic process.

$$\begin{aligned} \frac{\delta V}{\delta t} + \bigtriangledown \cdot flux &= 0 \\ \frac{\delta V}{\delta t} &= -\bigtriangledown \cdot flux \\ &= \bigtriangledown \cdot (D \bigtriangledown V) \end{aligned} \tag{4.1}$$

Practical application of Eqn 4.1 demands a result to be extracted at some $t > 0$, where the result at $t = 0$ is the original volume. No general analytical solution exists and so a consistent and discrete Euler-forward solution takes a Taylor's series approximation for the evolution truncated at the first-order term (the terms shown in Eqn 4.2) and defines a time-step $\Delta\tau$ to advance t after each iteration.

$$V^{t+\Delta\tau} = V^t + \Delta\tau \cdot \frac{\delta}{\delta t} V^t + ... \tag{4.2}$$

The Courant-Friedrichs-Lewy (CFL) condition [Courant et al., 1928] ensures that this numerical approximation of diffusion (Eqn 4.2) is at most as fast as the differential equation (Eqn 4.1) by enforcing an upper bound on $\Delta\tau$. This bound guarantees stability – defining a good model of evolution by limiting the amount of change that can occur in each iteration – when solving partial differential equations using explicit time-marching schemes. The Lax-Richtmyer Equivalence Theorem states that a consistent and stable approximation can be considered convergent [Strikwerda, 1989], and so the accuracy of the numerical approximation can be altered: as $\Delta\tau \to 0$ then the approximation will improve (at the penalty of longer processing time), and the opposite will occur for $\Delta\tau$ approaching the CFL bound, with error $O([\Delta\tau]^2)$. The CFL bound for diffusion in multi-channel volumetric data derived in [Gerig et al., 1992] is reproduced in Dfn 4.1.1for future reference.

Definition 4.1.1 (CFL bound for diffusion in multi-channel volumetric data) Diffusion can occur in R directions with the distance to a neighbouring voxel i in a direction chosen from R denoted as Δr_i. This diffusion is stable as long as the weight of the central element $\overline{V}_0$ exceeds the weights of its neighbouring elements $\overline{V}_i$, which are the first and second groups of non-V terms respectively in the second line of Eqn 4.3. g is a spatially-adaptive diffusion co-efficient replacing the static diffusivity bias D from Eqn 4.1 to allow the desired local smoothing directions

and amplitudes to be described, and the amount of flow allowed to each neighbour is chosen to be inversely proportional to the square of Δr_i.

$$\begin{aligned} \overline{V}^{t+\Delta\tau} &\approx \overline{V}_0^t + \Delta\tau \sum_{i=1}^{R} \frac{1}{(\Delta r_i)^2} g_i (\overline{V}_i^t - \overline{V}_0^t) \\ &= \overline{V}_0^t \left(1 - \Delta\tau \sum_{i=1}^{R} \frac{1}{(\Delta r_i)^2} g_i \right) + \Delta\tau \sum_{i=1}^{R} \frac{1}{(\Delta r_i)^2} c_i \overline{V}_i^t \end{aligned} \tag{4.3}$$

An inequality can be formed from Eqn 4.3 to ensure stability by monotonic variation of the voxel intensities. In this CFL bound for $\Delta\tau$ (Eqn 4.4) the neighbour indexed by m is examined which has the smallest distance from the central element (and thus the most flow is allowed to it) that is assumed to be of unit magnitude. Gerig *et al.* considered the extreme case of maximum blurring ($k \to \infty$) which led to g approaching 1 – where the most diffusion is allowed to occur – as the limiting case for the bound.

$$\begin{aligned} \left(1 - \Delta\tau \sum_{i=1}^{R} \frac{1}{(\Delta r_i)^2} g_i \right) &\geq \frac{\Delta\tau g_m}{(\Delta r_m)^2} \\ 1 &\geq \Delta\tau \left(\frac{g_m}{(\Delta r_m)^2} + \sum_{i=1}^{R} \frac{1}{(\Delta r_i)^2} g_i \right) \\ \Delta\tau &\leq \frac{1}{\frac{g_m}{(\Delta r_m)^2} + \sum_{i=1}^{R} \frac{1}{(\Delta r_i)^2} g_i} \\ \Delta\tau &\leq \frac{1}{1 + \sum_{i=1}^{R} \frac{1}{(\Delta r_i)^2}} \end{aligned} \tag{4.4}$$

The design of the adaptive co-efficient g (used in Dfn 4.1.1) used in place of D (used in Eqn 4.1) controls the non-linearity of response to the local gradient; one of the first such co-efficients was employed in [Perona and Malik, 1990] (Eqn 4.5) to encourage intra-region smoothing (causing $g \to 1$) when the local gradient was low ($|\bigtriangledown V| \to 0$) and to inhibit inter-region smoothing ($g \to 0$) to preserve their edges ($|\bigtriangledown V| \to \infty$). If the gradient can be calculated at several scale-space levels, then this process can be appreciated as an edge detector examining zero-crossings of the second

derivative at many scales.

$$V^{t+\Delta\tau} = V^t + \Delta\tau \bigtriangledown \cdot [g(|\bigtriangledown V^t|) \bigtriangledown V^t] \tag{4.5}$$

The form of g is flexible: for example, Perona and Malik suggested rapidly-decreasing Lorentzian (Eqn 4.6) and Gaussian (Eqn 4.7) functions which both tend towards zero with increasing gradient (as illustrated in Fig 4.2), but produce higher-contrast edges and wider regions respectively. α controls the steepness of the co-efficient response, and usually ranges between 1 and 3. k also controls the rate at which g decreases as the gradient magnitude increases, and it is usually set proportional to the expected gradient strength of edges that need to be preserved (higher than those associated with noise).

$$g_L(|\bigtriangledown V|) = \frac{1}{1 + \left(\frac{|\bigtriangledown V|}{k}\right)^{1+\alpha}} \tag{4.6}$$

$$g_G(|\bigtriangledown V|) = e^{-\left(\frac{|\bigtriangledown V|}{k}\right)^{1+\alpha}} \tag{4.7}$$

In order to reduce sensitivity to the noise-induced gradient oscillations in g, smoothing of the gradient was popularised in [Catte et al., 1992]. Scale-specific smoothing was performed using a Gaussian filter G (parameterised by its standard deviation σ in Eqn 4.8) so that the maximum principle was preserved to promote the stability and uniqueness of the results.

$$V^{t+\Delta\tau} = V^t + \Delta\tau \bigtriangledown \cdot [g(|\bigtriangledown G_\sigma \otimes V^t|) \bigtriangledown V^t] \tag{4.8}$$

V will degenerate to a constant value as $t \to \infty$ using forward-time diffusion since the diffusion co-efficient g always produces a positive result, so the result is recovered at some $x \ll \infty$, where V^x has been smoothed to some extent. However, the form of g can be changed so that edge-sharpening responses can be induced from the local gradient in addition to the edge-preserving and region-smoothing responses of forward anisotropic diffusion. PVE reduction and edge enhancement demand not just edge preservation, where diffusion is halted across edges but can proceed normal to them [Black et al., 1998], but also changes to the sign of flow to allowing sharpening. One of the first such implementations of edge enhancement was Forward-and-Backward diffusion, proposed in [Gilboa et al., 2002]. Diffusion was halted at very high gradient edges (determined by a constant w in Eqn 4.9), forward diffusion occurred with positive g to smooth low gradient regions (determined by k), and backward diffusion using

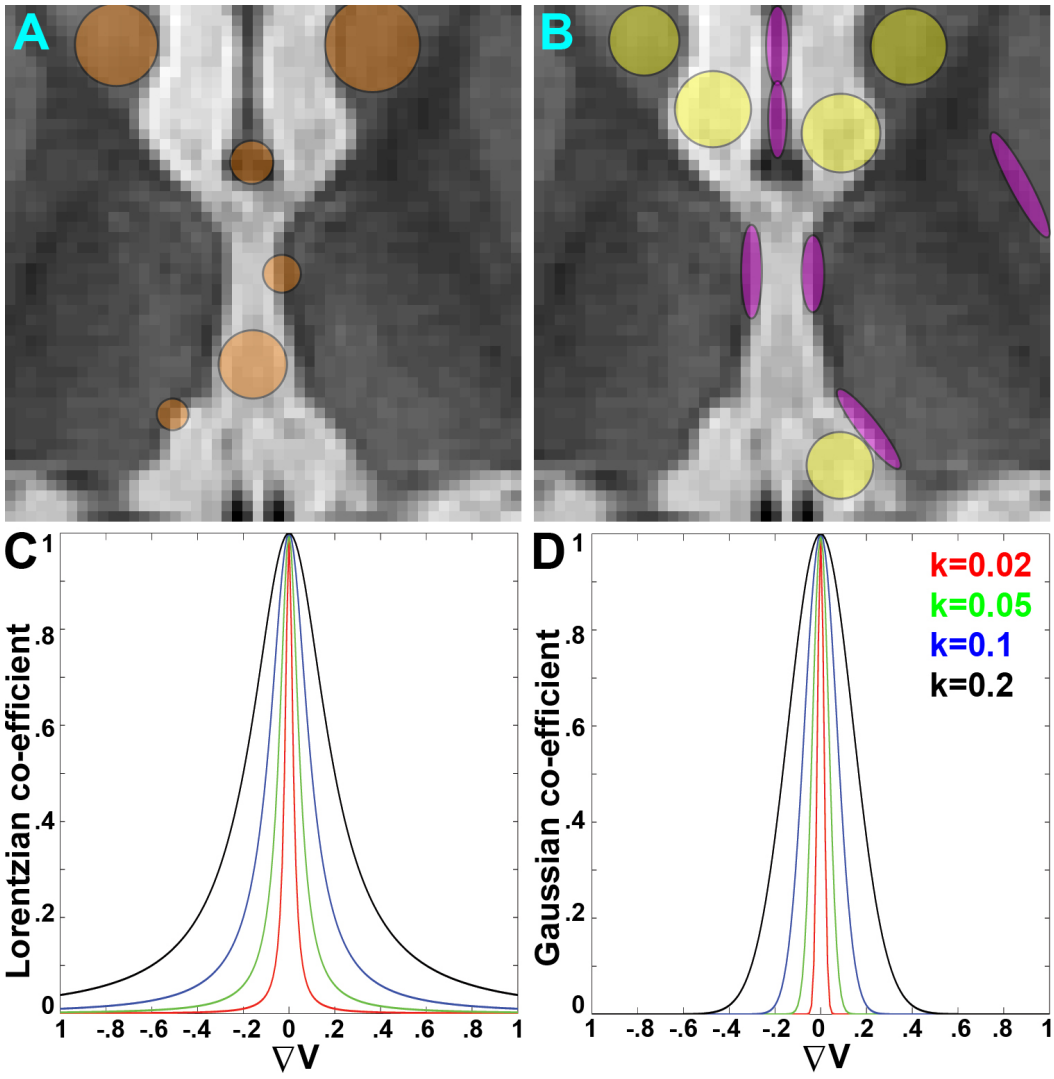

Figure 4.2: **Non-linear anisotropic diffusion**. PD data from ROI 6 is shown in **A** and **B**. **A**) Non-linear diffusion: diffusion is reduced across boundaries (small circles), but not restricted in low gradient regions (large circles). **B**) Anisotropic diffusion: diffusion is more anisotropic along high gradient edges (ellipses) but is more isotropic in low gradient regions (circles). **C**) Lorentzian-style diffusivity function (Eqn 4.6) with a range of values of k. **D**) Gaussian-style diffusivity function (Eqn 4.7) with a range of values of k. $\bigtriangledown$**V**: intensity gradient.

negative g sharpened other edges (determined by k_b) by promoting flow with relative strength γ against the local gradient.

$$g_{FAB}(|\bigtriangledown V|) = \frac{1}{1+(\frac{|\bigtriangledown V|}{k})^4} - \frac{\gamma}{1+(\frac{|\bigtriangledown V|-k_b}{w})^4} \tag{4.9}$$

Resolution enhancement of volumes using a similar diffusion co-efficient which is capable of enhancing edges is a candidate for performing both PVE reduction and homogeneous region smoothing (using smoothed gradients to provide some robustness to noise) when accompanied by an interpolation method to increase the resolution of the structural MRI volume data. Its computational efficiency and accuracy are determined by $\Delta\tau$ and it can be shown to give stable convergence when $\Delta\tau$ is set within the CFL bound. However, the fidelity constraints for such a choice of g need to be carefully considered since noise with a strong magnitude compared to the contrast between certain tissues – which is not uncommon in MRI – may be preserved and even enhanced, leading to degenerate results. These constraints were considered in [Salvado et al., 2006], forming the foundation of the proposed methods in this chapter.

4.1.2 Interpolation with reverse diffusion

The PVE reduction method presented in [Salvado et al., 2006] investigated effectively reversing the progression of time in diffusion, to produce a reverse diffusion process for improving visualisation of poorly-resolved blood vessels in structural MRI of the head and neck. A core assumption made was that PVE was caused by a sampling rate that did not capture sharp edges between tissue regions; this can be validated by visual inspection of the young brain volumes acquired at different resolutions (smoother and sharper borders of structures marked with red arrows are apparent in the bottom row of Fig 4.3). When voxels on tissue region boundaries were split into child sub-voxels with identical intensities by nearest-neighbour interpolation, reverse diffusion (illustrated in Fig 4.4A) could drive intensity between sub-voxels within the confines of each parent voxel's boundaries in the direction the higher intensity neighbouring regions.

The key contribution of this work was the development of fidelity constraints appropriate to MRI data and the mixel model. To enforce adiabatic flow constraints as well as to conserve signal inside voxels, flow of intensity to or from a sub-voxel indexed by i across the original voxel boundaries can be blocked by a binary boundary map M (Eqn 4.10). M is defined for each flow direction $r \in R$ and L is the upscaling ratio applied during nearest neighbour interpolation to create the initial sub-voxel

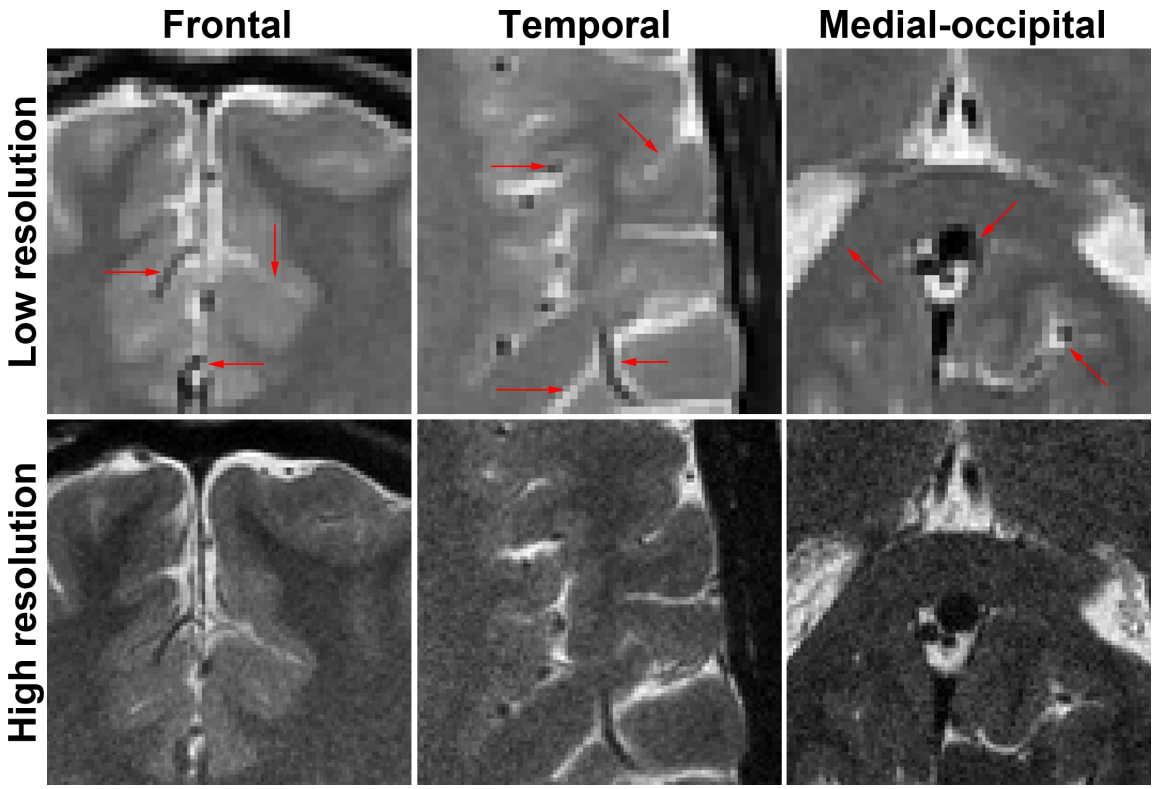

Figure 4.3: **Comparison of the same brain volume data acquired at two different resolutions**. **Top row**: Examples of T_2 data acquired with a $1 \times 1 \times 2$ mm^3 voxel size. **Bottom row**: T_2 data acquired with a $0.5 \times 0.5 \times 1$ mm^3 voxel size, formed from multiple averages as described in Section 2.2. Slight differences between the low and high resolution data are caused by changes in noise characteristics, acquisition parameters and slice coverage – in particular, the high resolution slices are only half as thick and so the low resolution slices incorporate contributions from additional tissue at each voxel. Red arrows are referenced in Section 4.1.2.

intensities.

$$M_{(i,r)} = \begin{cases} 0 & \text{if any}(i \bmod L_r = 0) \\ 1 & \text{otherwise.} \end{cases} \tag{4.10}$$

Reverse diffusion effectively operates with a negative time-step ($-\Delta\tau$); without special constraints, such a process would violate the inequality in Eqn 4.4 and thus lead to degeneration by the creation of massive unrestricted peaks and troughs in the intensities[5]. Flow limits based upon the capacity of the immediate neighbourhood are therefore required to prevent such expansion of intensity extrema but still allow PVE reduction (defined in Dfn 4.1.2). This fidelity constraint relies upon an implication in the stability of the continuity equation (Eqn 4.1) that there must be sufficient intensity that is allowed to be transported from a lower-intensity sub-voxel as well as sufficient capability in the higher-intensity sub-voxel to be able to receive it. Both these properties are measured in a noise-resistant manner using local-level-limited flow: intensity is not able to fall below its local low-level nor rise above its local high-level. These values are obtained through consideration of the intensity-ranked 26-connected neighbourhood. More extreme values of the low and high rank indices q_{min} and q_{max} from the median increased the sensitivity to noise but allowed steeper edges to be produced, as well as faster convergence towards a state where little flow occurred between iterations (signalling that iterations should terminate).

Definition 4.1.2 (Local level-limited flow) The **local low level difference** Q_i^{min} in the ascending rank-ordered neighbourhood Nh of the sub-voxel at position i is a non-negative value, determined by the difference in intensity of the sub-voxel indexed by i with that of the $(\text{median} - q_{min})^{th}$ ranked sub-voxel, where $median = \text{ceil}\left(\frac{size(Nh)}{2}\right)$ and $0 < q_{min} < \text{floor}\left(\frac{size(Nh)}{2}\right)$.

$$Q_i^{min}(V, q_{min}) = \max[0, V_i - ord_{i' \in Nh}^{ind}(V_{i'})] \quad \text{s.t.} \quad ind = median - q_{min} \tag{4.11}$$

The **local high level difference** Q_i^{max} is determined by the difference in intensity of the $(\text{median} + q_{max})^{th}$ ranked sub-voxel with that of the central sub-voxel, where $0 < q_{max} < \text{floor}\left(\frac{size(Nh)}{2}\right)$.

[5]This can be envisaged as the opposite situation to forward-time diffusion at $t = \infty$.

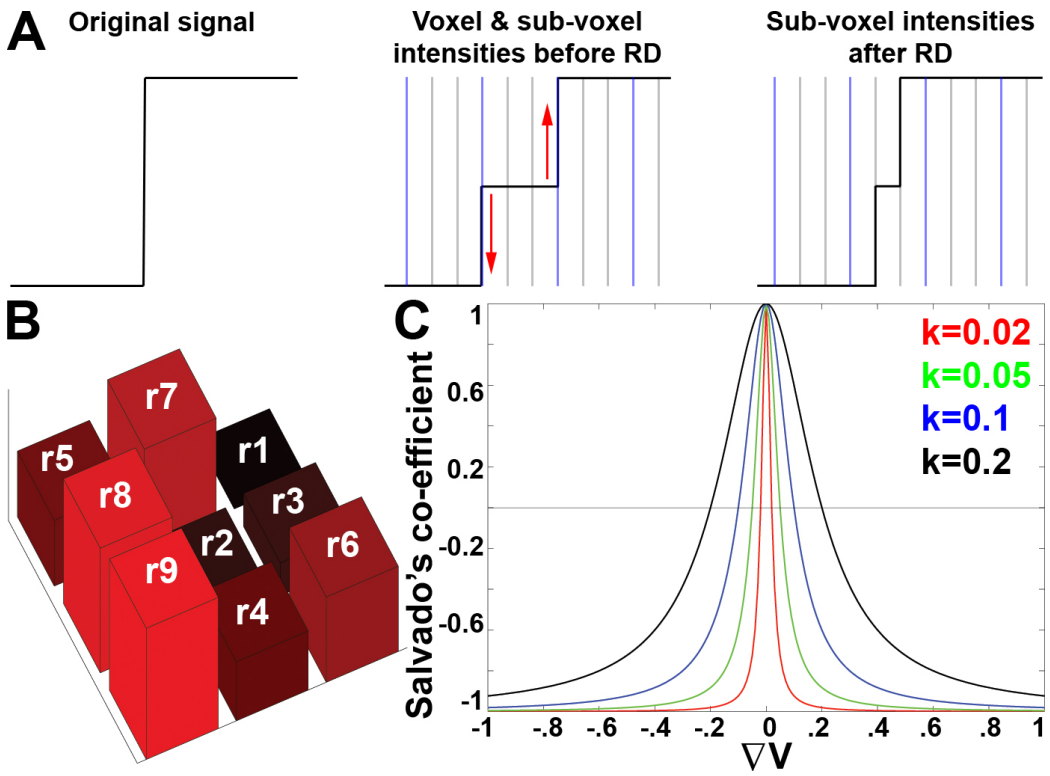

Figure 4.4: **Reverse diffusion processes**. **A**) Effect of reverse diffusion in 1D. **Left**: A step edge signal (black line) is apparent at infinite resolution. **Middle**: Discretisation leads to PVE due to undersampling, and PVE is apparent at both the normal resolution (blue lines) and the $L = 3\times$ nearest-neighbour upscaled level (gray lines). Arrows indicate the intensity changes in sub-voxels during reverse diffusion. **Right**: Reverse diffusion is able to restore the step edge by a degree proportional to L, with respect to neighbouring minimum and maximum intensity values as well as the normal resolution boundaries. **B**) Rank ordering of intensities in a 2D neighbourhood. **r1**, the darkest shade of red, denotes the lowest ranking pixel with the lowest intensity; conversely, **r9** denotes the highest ranking pixel with the highest intensity and is the fullest red. **C**) Response of the diffusion co-efficient of reverse anisotropic diffusion (Eqn 4.17). $\bigtriangledown$**V**: intensity gradient.

$$Q_i^{max}(V, q_{max}) = \max[0, ord_{i' \in Nh}^{ind}(V_{i'}) - V_i] \quad \text{s.t.} \quad ind = median + q_{max} \quad (4.12)$$

The **local level-limited flow** $Q_{i,i'}$ between the sub-voxel indexed by i and a neighbouring sub-voxel indexed by i' constrains intensity flow F such that it does not cause the local high level difference to be exceeded, nor cause a fall by more than the local low level difference. The sign of the unsmoothed gradient of V determines where intensity is flowing from and to, and thus which local levels should be respected. The parameters of F, Q^{min} and Q^{max} are not noted in Eqn 4.13 in the interests of saving space.

$$Q_{i,i'}(V, q_{min}, q_{max}, \sigma) = \begin{cases} \min[F_i, \min[Q_{i'}^{min}, Q_i^{max}]] & \text{if } V_{i'} - V_i < 0 \\ \max[F_i, -\min[Q_{i'}^{max}, Q_i^{min}]] & \text{if } V_{i'} - V_i \geq 0 \end{cases} \quad (4.13)$$

Combining the local level limited flow (Dfn 4.1.2) with the voxel boundary flow constraints (Eqn 4.10) allows the discrete reverse diffusion scheme to be completely specified (Eqn 4.14). The unconstrained flow F was initially a linear function of the smoothed gradient (Eqn 4.15) but in a subsequent paper [Salvado and Wilson, 2006] non-linearity was implemented using an adaptive diffusion co-efficient g (Eqn 4.16) to create a reverse anisotropic diffusion (RAD) process following an initial reverse diffusion (RD) phase.

$$V_i^{t+\Delta\tau} = V_i^t + \Delta\tau . \sum_{i' \in Nh(i)} M_{i'} Q_{i,i'}(V^t, q_{min}, q_{max}, F) \quad (4.14)$$

$$F_{RD}(V, \sigma) = -(\bigtriangledown G_\sigma \otimes V) \quad (4.15)$$

$$F_{RAD}(V, \sigma) = -g_{RAD}(|\bigtriangledown G_\sigma \otimes V|).(\bigtriangledown G_\sigma \otimes V) \quad (4.16)$$

The two phases each had a distinct role. The RD phase caused most of the edge enhancement and PVE reduction, while noise and low frequency variation inside region borders was smoothed and edges were slightly further sharpened in the RAD phase. The latter phase demanded the removal of voxel boundary flow constraints to implement inter-voxel smoothing, and so M was redefined to only enforce adiabatic

flow constraints using the volume boundary indices Ω. The original volume cannot be exactly recovered by downsampling after the RAD phase and so only a small number of RAD iterations were performed to minimise the opportunity for any substantial changes to occur.

There are similarities between the diffusion co-efficient g (Eqn 4.17) used in the RAD phase and the one designed for forward-and-backward diffusion [Gilboa et al., 2002]: each has a forward and a backward diffusion term, with the latter given a weight relative to the former. Salvado *et al.*'s g (Eqn 4.17) is simpler in concept since there is only a single parameter k and each term is a Lorentzian function. The forward time diffusion term is dominant when the gradient u is small compared to k and it encourages smoothing in homogeneous low gradient areas ($g(u) \rightarrow 1$ as $u \rightarrow 0$). The second reverse time diffusion term is strongest when u is large compared to k and it promotes sharpening at tissue boundaries ($g(u) \rightarrow -\alpha$ as $u \rightarrow \infty$). The magnitude of $0 \leq \alpha \leq 1$ describes the amount of sharpening that can occur in the reverse time diffusion term. Since the CFL bound is supplemented with the flow constraints (Dfn 4.1.2), then this kind of diffusion is made stable and g can safely take values in the range $[-1, 1]$.

$$g_{RAD}(u) = \frac{1}{1 + (\frac{|u|}{k})^2} + \alpha. \left(\frac{1}{1 + (\frac{|u|}{k})^2} - 1 \right) \tag{4.17}$$

The typical behaviour of the absolute flow between iterations, defined as $\sum_{i \in voxels} |V_i^t - V_i^{t-\Delta\tau}|$, is convergent towards zero as $t \rightarrow \infty$ since the intensity gradients will become either extremely low or high in magnitude at high t. However, running the scheme until very large t is reached is not computationally viable and a more moderate choice of t will likely provide a result which has not been over-smoothed nor over-sharpened. The following section details the modifications to this diffusion scheme necessary for application to dual-channel structural MRI of the brain with poorly-resolved, contrasting thin structures for which the pure tissue intensities should be restored at the sub-voxel level.

4.2 Modifications to interpolation with reverse diffusion

In [Caselles et al., 1998] several axioms were outlined that should be satisfied by interpolation operators. These are incorporated below in a list of requirements for a new resolution enhancement process applicable to dual-channel MRI data containing thin

contrasting structures:

1. The computation must be convergent (as described in Section 4.1.1) and processing on a volume should complete in the order of hours, rather than days;

2. The method should not be too sensitive to internal parameters that may vary from volume to volume, such as the slice thickness, contrast or noise level;

3. Texture is not an important feature and so variation of intensity in homogeneous tissue regions can therefore be attributed to noise (presumed additive Gaussian, as investigated in Section 5.2.2) which is uncorrelated between the acquisitions of volumes with different image weighting factors and should be eliminated;

4. Gradient calculations should use all the available data channels, to exploit the different tissue contrasts as well as their uncorrelated noise;

5. The mean intensity, shape and size of anatomical structures as well as the location of their borders should not be altered, though reduction of PVE and intensity restoration near edges will manifest in slight changes at the sub-voxel scale;

6. Tissue region boundaries affected by PVE should be made more crisp by the formation of sub-voxels with intensities similar to those of the regions either side of the boundary;

7. Extra features should not be generated and noise should not be enhanced, and the preservation as well as restoration of poorly-resolved thin contrasting structures should be tackled.

With respect to these objectives, the original resolution enhancement methods [Salvado et al., 2006][Salvado and Wilson, 2006] described in the previous subsection are not completely suitable for the task. Most importantly, there is no mechanism to allow the core intensities of very poorly-resolved thin structures to be restored (requirement 7) since the extreme local level differences will not manifest properly due to PVE, especially with robust choices of rank indices q_{min} and q_{max}. This failure causes reduced flow and thus less edge enhancement to occur in the RD phase in the region of thin structures, but a more serious effect in the RAD phase is that these structures (marked with red arrows in Fig 4.5) can be blurred and reduced in contrast without careful consideration of k (used in Eqn 4.17) and the standard deviation of the Gaussian smoothing of the gradients (causing concerns for fulfilling requirements 5-7).

Other issues that need to be resolved include there being no description of methods to process two or more co-registered data channels simultaneously while ensuring consistency of their structures' boundaries (requirement 5, with notable discrepancies in boundary locations between the separately-processed image weighting factors marked with blue arrows in Fig 4.5), and also the scale space level at which Gaussian filtering (Eqn 4.8) of the gradient is performed may cause insufficient sensitivity to thin structures (requirement 7). Finally, more diffusion directions may be considered in addition to those along the main data axes in order to reduce the number of iterations required to reach convergence (tackling requirement 1) and to allow more direct reduction of PVE to be accomplished regardless of the boundary orientation.

In the thin structure detection process, binary maps B (Dfn 3.4.5) defining the likely presence and optimum scale of thin structures were produced, and ϕ (defined in Section 3.4.4) denoted the index of the shaped and oriented filter with the closest principal normal direction for each sub-voxel. The volume upsizing ratio L used in this chapter's work should match that used during thin structure detection (Chapter 3); this was kept to small integer values due to its direct multiplicative effect on the data size, resulting in increased memory usage and processing time. The following sections introduce dual-channel gradient calculation (Section 4.2.1), 26-connected flow (Section 4.2.2) and thin structure sub-voxel intensity restoration using B and ϕ (Section 4.2.3) to help tackle the outstanding issues listed above. A final summary of the complete volume resolution enhancement process is listed in Section 4.2.5.

4.2.1 Dual-channel gradient calculation

The benefit of studying a vector of co-registered values for each MRI weighting factor at every voxel is that the robustness of evaluating the local gradient in the face of noise – which is uncorrelated between the modalities – may be improved. Combinations of different tissue contrasts can also provide more sensitive structure boundary detection for tissues which are poorly contrasted in one image weighting factor (such as BG/WM in T_2 and CSF/GM in PD). Due to these differences in contrast, performing anisotropic diffusion smoothing on the image weighting factors independently could cause inconsistencies in structure boundary locations (such as the WM/GM boundaries marked with arrows in Fig 4.6) and so using multi-channel gradients to calculate the diffusion co-efficient can help to ensure consistent edge localisation.

The Euclidean norm of the gradients from both volumes (Dfn 4.2.1) was employed

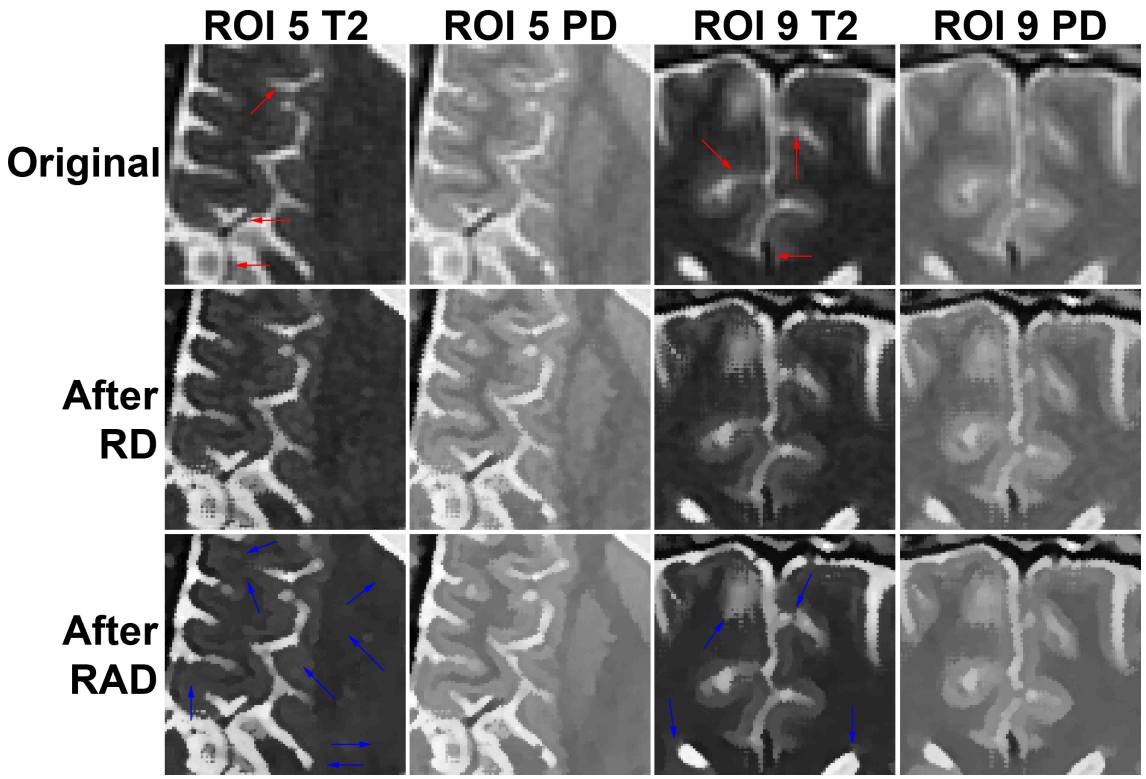

Figure 4.5: **Poor flow in the region of thin structures**. **Top row**: original T_2 and PD data from ROIs from the LBC1936 dataset. **Middle row**: enhancement of the regions in the top row after 50 iterations of reverse diffusion using the method presented in [Salvado et al., 2006]. The parameters used were $q_{min} = q_{max} = 6$, $k = 0.01$, $\Delta\tau = \frac{1}{7}$. **Bottom row**: enhancement of the regions in the middle row after 15 iterations of reverse anisotropic diffusion [Salvado and Wilson, 2006]. The same parameters were used. **RD**: reverse diffusion. **RAD**: reverse anisotropic diffusion. Coloured arrows are referenced in Section 4.2.

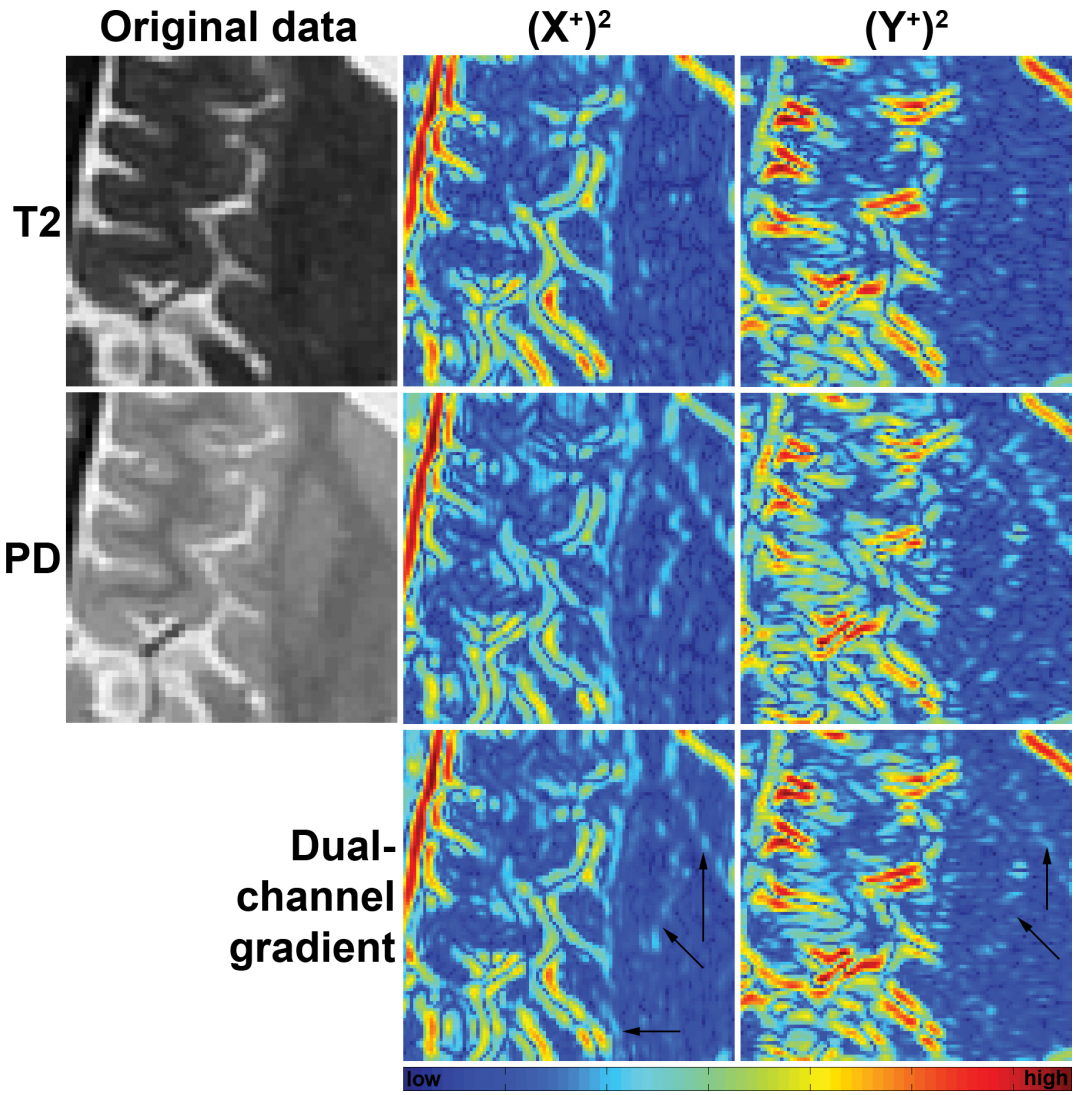

Figure 4.6: **Comparison of dual-channel gradients with single-channel gradients**. **First row**: T_2-weighted data for ROI 5 and gradients using this image weighting factor's data only. **Second row**: PD-weighted data for ROI 5 and gradients using this image weighting factor's data only. **Third row**: Gradients using both image weighting factors' data. **Second and third columns**: Gradients for orientations X^+ (positive horizontal axis direction) and Y^+ (positive vertical axis direction) respectively. Black arrows are referenced in Section 4.2.1. The square of the gradient is presented so that the magnitude can be examined and also because low-gradient fluctuations in pure tissue regions will be suppressed.

in the calculation of the diffusion co-efficient (Eqn 4.15), as proposed in [Gerig et al., 1992]. Since squared gradient magnitudes are employed then their signs are lost – however the absolute operator is used in the diffusion co-efficient (Eqn 4.17), and so the sign is irrelevant.

Definition 4.2.1 The **smoothed dual-channel gradient** $\bigtriangledown_\sigma \overline{V}$ of a co-registered T_2-PD volume pair $\overline{V}$ employs the Euclidean norm of the individual smoothed gradients. G_σ is a Gaussian filter with standard deviation σ.

$$\bigtriangledown_\sigma \overline{V} \;=\; \sqrt{(\bigtriangledown[G(\sigma)\otimes V_{T2}])^2 + (\bigtriangledown[G(\sigma)\otimes V_{PD}])^2} \qquad (4.18)$$

4.2.2 Computation of time-step bounds and 26-connected flow

The literal interpretation of the divergence operator, used in Eqn 4.1, as $\bigtriangledown \cdot V = \frac{\delta V_x}{\delta x} + \frac{\delta V_y}{\delta y} + \frac{\delta V_z}{\delta z}$ prompts most implementations of diffusion-based smoothing on volumes to apply only 6-connected flow (in both orientations along each main axis). However, the number of flow directions can be extended to cover all 26 adjacent voxels – up to third-order neighbours as listed in Table 7.1. The main benefit of this change is that the dependence on flow through intermediate sub-voxels (first- or second-order neighbours) is removed and so PVE at boundaries of any orientation will be more quickly and directly reduced. The chief disadvantage of using 20 more flow orientations lies in a more than three-fold increase in the storage and computation time required for calculations involving the gradient at each iteration, though the potential improvements to PVE reduction are thought to justify this cost.

Convergence of reverse diffusion must remain stable after any changes are made to the scheme, but upon closer examination of the original definition of the local level-limited flow $Q_{i,i'}$ (Dfn 4.1.2) proposed in [Salvado et al., 2006], this was not guaranteed. The sign of the unsmoothed gradient $\bigtriangledown V$ was used to determine which local levels should form the bounds – in terms of which sub-voxel the intensity is flowing from, setting the low level, and where it is flowing to, affecting the high level – but it will not always be equal to that of the filtered gradient $\bigtriangledown(G_\sigma \otimes V)$ which determines the flow (used in Eqn 4.15). Therefore, flow can actually surpass the local low and high level differences; to remedy this situation then the sign of the filtered gradient

must determine which levels form the bounds. $Q_{i,i'}$ is redefined as such in Dfn 4.2.2 so that stable convergence is assured.

Definition 4.2.2 (New local level-limited flow) The **new local level-limited flow** $Q_{i,i'}$ from sub-voxel index i to neighbouring sub-voxel i', extending Dfn 4.1.2, is constrained to cap intensity flow F between them such that it does not cause the local high level difference to be exceeded, nor to cause a fall by more than the local low level difference. The smoothed gradient $\bigtriangledown_\sigma \overline{V}_i$ (Dfn 4.2.1) now determines which local levels are checked. The parameters of F, Q^{min} and Q^{max} are not noted below and the parameters of $Q_{i,i'}(\overline{V}, q_{min}, q_{max}, \sigma)$ are abbreviated to $(...)$ in the interests of saving space.

$$Q_{i,i'}(...) = \begin{cases} \min[F_i, \min[Q_{i'}^{min}, Q_i^{max}]] & \text{if } \bigtriangledown_\sigma \overline{V}_i < 0 \\ \max[F_i, -\min[Q_{i'}^{max}, Q_i^{min}]] & \text{if } \bigtriangledown_\sigma \overline{V}_i \geq 0 \end{cases} \tag{4.19}$$

In the RD phase a re-analysis of the CFL bound for the diffusion time-step $\Delta\tau$ (Eqn 4.4) revealed that $\Delta\tau$ can be increased (and in turn the allowed speed of diffusion) when the number of diffusion directions is limited locally by voxel boundary constraints M: no flow will occur to or from sub-voxels next to boundaries in directions passing across those boundaries. These directions can therefore be removed from the direction set R in the denominator of the CFL bound. In order that the bound for each sub-voxel remains consistent with those of its neighbours – and recursively with the entire volume – then the volume-wide minimum time step value must be chosen (as detailed in Table 7.2).

The most restrictive $\Delta\tau$ (the last row of Table 7.2) must be used when flow can occur in any direction in the RAD phase – when flow is not constrained by any adjacent voxel boundaries – or when all elements of the upscaling ratio L are ≥ 3 in the RD phase. When any element of L is 2 or 1 in the RD phase, then the flow directions become restricted and $\Delta\tau$ can be relaxed.

4.2.3 Preservation of thin structures and restoration of their intensities

Reduced flow in the region of thin structures (marked with red arrows in Fig 4.5) using the original resolution enhancement methods can be attributed to several factors.

Firstly, their poor resolution leads to the formation of few thin structure voxels with close-to-pure tissue intensities, as observed in Fig 4.7 (black arrows): it can be seen that most sulcal voxels have less extreme local high levels than the lateral ventricle located in the top-right corner. The local level difference (Dfn 4.1.2) rank indices will therefore need to be more extreme in these regions in order to potentially capture any (at least close to) pure tissue intensities. To prevent these structures' sparse extreme intensities from being treated as noise by the rank order indices, a rank order modifier *qn* can supplement the existing rank order indices q_{min} and q_{max}. Using *qn* also endows a greater sensitivity to noise and so it is only applied to the tissue-specific extremity – the high difference Q^{max} for bright CSF, and the low difference Q^{min} for dark BG – and only in the particular thin structure regions identified in B (Dfn 3.4.5).

In Section 3.2 anisotropic filters which are similarly shaped to the local structure were shown to help preserve shape and intensity contrast during smoothing. The effect of applying these filters in thin contrasting regions will be that sub-voxels with contrasting intensities will take more high filter values inside thin structures than if isotropic filters were used, which will produce a more extreme response. Therefore the smoothed gradient $\bigtriangledown_\sigma \overline{V}$ (used in Eqns 4.15 and 4.16) will instead be smoothed by the appropriate anisotropic filter (specified by ϕ in Section 3.4.4) in thin structure regions in order to increase the gradient magnitude and lead to more rapid reduction of PVE, and by an isotropic filter in other regions as normal.

In the event that there are few or even no voxels with close-to-pure tissue intensities in thin structure regions, then the appropriate local level will be less extreme and pure tissue intensity sub-voxels will never be formed by local-level-limited diffusion; intensity restoration processes are then needed to effectively enhance the resolution. In Salvado's PhD thesis [Salvado, 2006] he observed that given accurate *a priori* information about the location and true magnitude of local intensity extrema, those which had been reduced in magnitude by insufficient sampling could be recovered by setting the appropriate local levels to representative intensities during diffusion. For this work I^{BG} and I^{CSF} are used to calculate the local level differences of flagged thin structure sub-voxels in B, and class intensity means computed by a simple segmentation method can be used to estimate I. For more robust estimation of I then high membership (likely pure) samples of tissue class intensities can be selected and a morphological erosion step – similar in form to that proposed in Proc 3.4.2 – can extract samples from the cores of structures where pure tissue is more likely to be present.

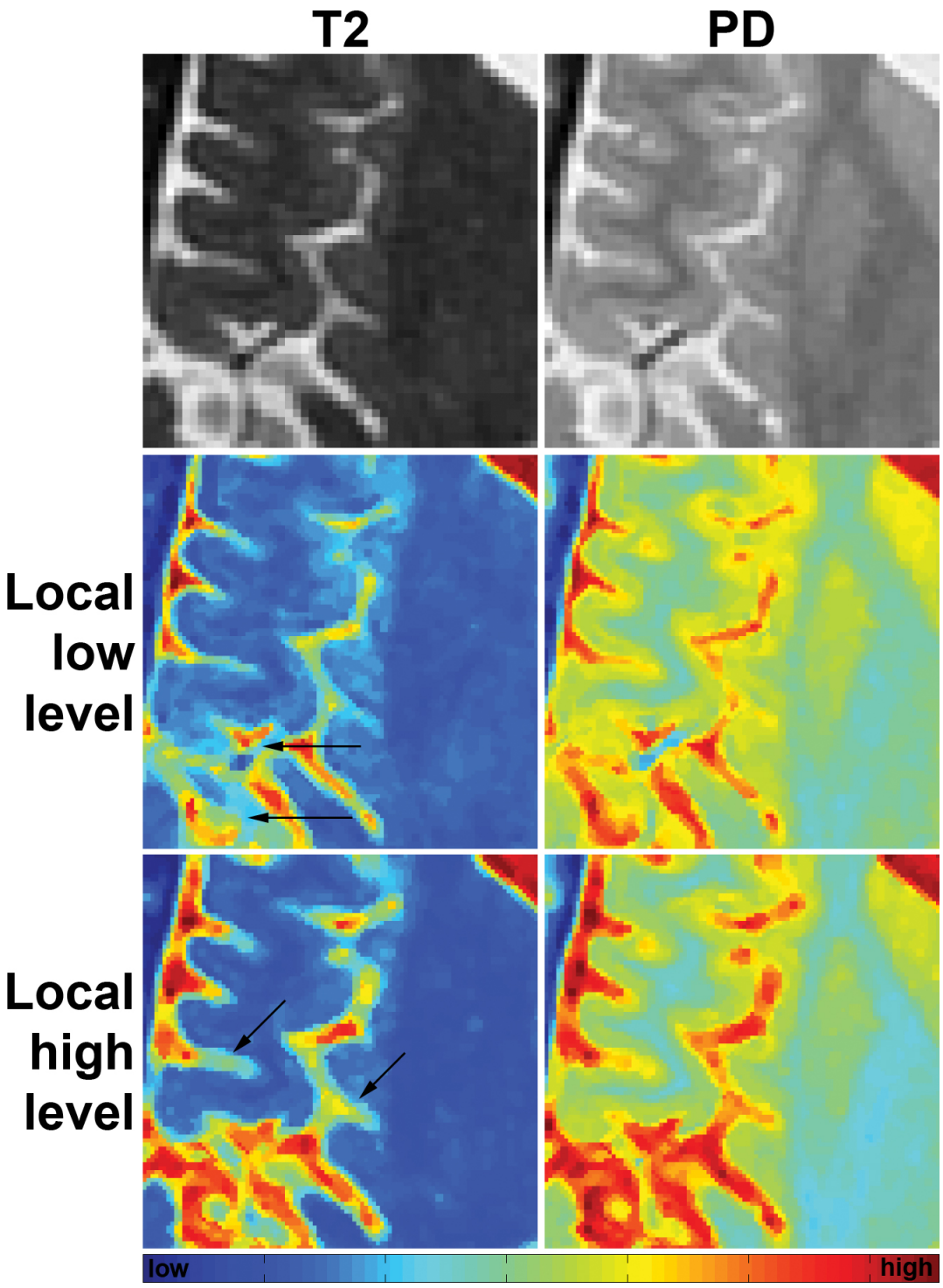

Figure 4.7: **Local levels for thin structures are less extreme than the pure tissue intensities at their cores**. **A** & **B**) T_2 and PD data from ROI 5. **C** & **D**) Local low levels for **A** and **B** ($ord^{\text{median}-7}$). **E** & **F**) Local high levels for **A** and **B** ($ord^{\text{median}+7}$). ord is the local neighbourhood intensity rank ordering function, as used in Dfn 4.2.3. Black arrows are referenced in Section 4.2.3.

Definition 4.2.3 (New local level differences) The **new local low level difference** Q_i^{min} supplements the original definition (Dfn 4.1.2) with both a predefined local low intensity I^{BG} and a more extreme $(\text{median} - q_{min} - qn)^{th}$ rank index for sub-voxels marked as BG structures in B. Consequently the limits of q_{min} must be redefined as $0 < q_{min} \leq \text{floor}\left(\frac{size(Nh)}{2}\right) - qn$. The parameters of $Q_i^{min}(\overline{V}, q_{min}, qn, B, I^{BG})$ are omitted in the interest of saving space. The low rank index $ind = \text{median}(Nh) - q_{min}$, and the more extreme low rank index $ind2 = ind - qn$.

$$Q_i^{min} = \begin{cases} \max[0, \overline{V}_i - \min(ord_{i' \in Nh}^{ind2}(\overline{V}_{i'}), I^{BG})] & \text{if } B_i^{BG}; \\ \max[0, \overline{V}_i - ord_{i' \in Nh}^{ind}(\overline{V}_{i'})] & \text{otherwise.} \end{cases} \tag{4.20}$$

The **new local high level difference** Q_i^{max} similarly supplements the original with a predefined local high intensity I^{CSF} as well as a more extreme $(\text{median} + q_{max} + qn)^{th}$ rank index for sub-voxels flagged in B^{CSF}. Again, the limits of q_{max} must be redefined as $0 < q_{max} \leq \text{floor}\left(\frac{size(Nh)}{2}\right) - qn$. The high rank index $ind = \text{median}(Nh) + q_{max}$, and the more extreme high rank index $ind2 = ind + qn$.

$$Q_i^{max} = \begin{cases} \max[0, \max(ord_{i' \in Nh}^{ind2}(\overline{V}_{i'}), I^{CSF}) - \overline{V}_i] & \text{if } B_i^{CSF}; \\ \max[0, ord_{i' \in Nh}^{ind}(\overline{V}_{i'}) - \overline{V}_i] & \text{otherwise.} \end{cases} \tag{4.21}$$

4.2.4 Termination criterion

In [Salvado et al., 2006] the observation was made that the total flow occurring each iteration of RD or RAD from $t = 0$ decreased monotonically[6], when trying to determine at which t an acceptable result had been produced. As $t \to \infty$ the gradient at tissue boundaries is sharpened, and in the RAD phase only the interiors of homogeneous regions become smoothed; these desired properties manifest in reduced relevant local level differences (Dfn 4.2.3) and thus a reduction in flow allowed. Following the reasoning of Salvado *et al.*'s observation then diffusion iterations were allowed to terminate once flow had fallen below a fraction of the maximum observed. The mean

[6]This has been demonstrated many times in Section 4.3.

absolute error was selected as a measure of the reduction in flow between iterations. It was modified in Dfn 4.2.4 to accommodate the dual-channel data and also to inspect all sequential time-point pairs in a window of time-points so that short-lived poor flow did not lead to termination.

Definition 4.2.4 (Multi-channel windowed mean Euclidean error) The **multi-channel windowed mean Euclidean error** assesses a window of time-points of length $x > 0$ (or equivalently, a number of iterations x) up to the current time (or iteration) t. $\|...\|_2$ denotes the Euclidean norm of the dual-channel volume intensity difference, $\overline{V}_i^y$ is the dual-channel volume intensity for sub-voxel i at time y, and $a:b$ summarises a sequence of time-points between times (or iterations) a and b with step $\Delta\tau$.

$$\text{MWMEE}(\overline{V}^{(t-x):t}) = \max_{y=(t-x+1):t} \left(\frac{\sum_{i \in voxels} \|\overline{V}_i^y - \overline{V}_i^{y-1}\|_2}{\#voxels} \right) \tag{4.22}$$

4.2.5 Processing steps and algorithm summaries

Procedure 4.2.1 (Pre-processing) Before the RD and RAD stages (Proc 4.2.2) can be performed on the dual-echo T_2-PD volumes, the following operations must take place:

1. The volumes must undergo skull stripping as described in Section 2.2 and the volume intensities should be normalised to a zero minimum and unit maximum;
2. The thin structure map B (upsized by ratio L), the shaped and oriented anisotropic filter bank, and the filter indices ϕ for each sub-voxel then need to be computed as described in Chapter 3;
3. FCM segmentation (described in Section 5.3.2) must be performed to extract likely pure tissue regions of BG and CSF, which are then morphologically

eroded. Their means form the representative pure tissue intensities I^{BG} and I^{CSF};

4. The normalised volumes are upsized by ratio L using nearest neighbour interpolation;

5. The voxel boundary flow limits M (Eqn 4.10) are computed;

6. Referring to Table 7.2, the time-step $\Delta\tau$ for the RD phase is set to the maximum possible value over all sub-voxels that does not exceed the CFL bound, given L and the volume's slice thickness h. The smallest $\Delta\tau$ (the final row of the table) is set for the RAD phase, as detailed in Section 4.2.2;

7. The expected noise level k of the volumes used in the calculation of the diffusion co-efficient is estimated using the median absolute deviation method [Black et al., 1998].

Procedure 4.2.2 (Reverse diffusion and reverse anisotropic diffusion) After pre-processing (Proc 4.2.1) the RD stage must be performed before the RAD stage, and the processing for both is similar:

1. The smoothed dual-channel gradient of the volumes is calculated (as described in Section 4.2.1) in 13 orientations using the anisotropic filter indexed by ϕ in regions flagged in B, and with an isotropic filter otherwise;

2. The flow is computed using a linear function of the gradient for the RD phase and a non-linear function (using a diffusion co-efficient) in the RAD phase, as described in Sections 4.2.1 and 4.2.2;

3. The local low and high level differences (Dfn 4.2.3) are then calculated to limit flow and thus ensure stable diffusion, using I^{BG} or I^{CSF} as the minimum intensity extremum for sub-voxels flagged in B;

4. The volumes are updated according to the diffusion time-step $\Delta\tau$ with the weighted flow in each direction (Eqn 4.14) constrained by the local level differences as well as the boundary constraints M;

5. The termination criterion is checked (as described in Section 4.2.4). If it has not been reached then processing resumes at step 2 with $t = t + \Delta\tau$, otherwise the resolution-enhanced volumes are returned.

4.3 Demonstrations

The initial demonstrations presented in this section compare the differences in performance between the individual modifications to the resolution enhancement methods (described in Section 4.2) with the original work [Salvado et al., 2006][Salvado and Wilson, 2006]. In concordance with the testing performed in Section 3.5, ROI 5 will mainly be studied due to its inclusion of several different anatomical structure types including blood vessels, cortical sulci, convoluted cortex, deep GM nuclei, the skull, the edge of the lateral ventricle and thin regions of WM. A similar region from the young brain dataset (previously used in Fig 3.17C) with poor WM/GM contrast in T_2 and poor GM/CSF contrast in PD has also been selected in order to examine the methods' robustness. Since a ground truth – effectively, a higher resolution volume of the same brain – is not available for either the young brain or LBC1936 datasets then the majority of observations in the following two subsections will be qualitative.

More quantitative assessment has been performed on the MRI data acquired at both typical clinical and high resolutions (in Section 4.3.3) where this type of assessment is made possible. In all the demonstrations in this section the parameters used to create the thin structure maps B (Dfn 3.4.5), the anisotropic shaped and oriented filter set and the filter indices ϕ are identical to those outlined in Section 3.5, and unless otherwise stated the following parameters have been applied to the remaining processes:

- The volume upsizing ratio $L = [2,2,2]$ matched that used during thin structure detection (Chapter 3);
- The tolerance of flow change for termination of resolution enhancement was set to 30% of the maximum MWMEE (Dfn 4.2.4) encountered for the RD phase and to 50% of the maximum for the RAD phase. These levels were chosen so that processing completed in a reasonable timeframe and to prevent excessively sharpened results from being produced, with the latter factor assessed by visual inspection;

- The rank order indices were set fairly high at $q_{min} = q_{max} = 7$ to allow sensitivity to edges but also to provide some robustness to noise, assessed by visual inspection of the results. In thin structure regions the additional rank order modifier $qn = 2$ was applied to either q_{min} or q_{max} to give a little extra sensitivity to the appropriate thin structure intensity extrema;

- The maximum possible value of the diffusion time-step $\Delta\tau$ was used (as described in Section 4.2.2) so as few iterations were performed as possible. Given the value of L, this was $\frac{1}{1+2+\frac{1}{2}+\frac{1}{h^2}+\frac{2}{1+h^2}+\frac{1}{2+h^2}}$ for the RD phase and $\frac{1}{1+4+\frac{4}{2}+\frac{2}{h^2}+\frac{8}{1+h^2}+\frac{8}{2+h^2}}$ for the RAD phase, where $h = 2$ is the slice thickness relative to the in-plane voxel dimensions for every dataset studied;

- FCM clustering (described in Section 5.3.2) was used to discover the representative intensities I of the BG and CSF tissue classes needed for intensity restoration of thin structures (detailed in Section 4.2.3), using the parameter values noted in Section 5.4.

4.3.1 Changes to flow and gradient calculations

Figs 4.8 and 4.9 compare the performance of resolution enhancement using the original 6-connected, single-channel gradient and flow calculations performed in [Salvado and Wilson, 2006][7] with the new 26-connected, dual-channel counterparts proposed in this chapter. No shaped and oriented filtering nor modification of the local levels (using q_n in Dfn 4.2.2) were used in thin structure regions, so that the changes could be assessed in isolation from the thin structure-specific processing. The most striking difference between the results is the smoother and more natural shape of structure edges apparent in the modified processing's results (marked with aqua arrows); this effect is made more clear in panels D and H of each figure where the intensity differences between the results have been localised to structure borders. Some tissue boundaries are also more consistently located between the image weighting factors (brown arrow) due to the consideration of the different contrasts in both data channels when using the new methods.

The contrast between tissues has not been changed from the original data – except at borders, where PVE is reduced – and so intensity extrema relevant to pure BG and

[7] However, they do incorporate the fix to the local level selection outlined in Section 4.2.2 to ensure degenerate results were not formed.

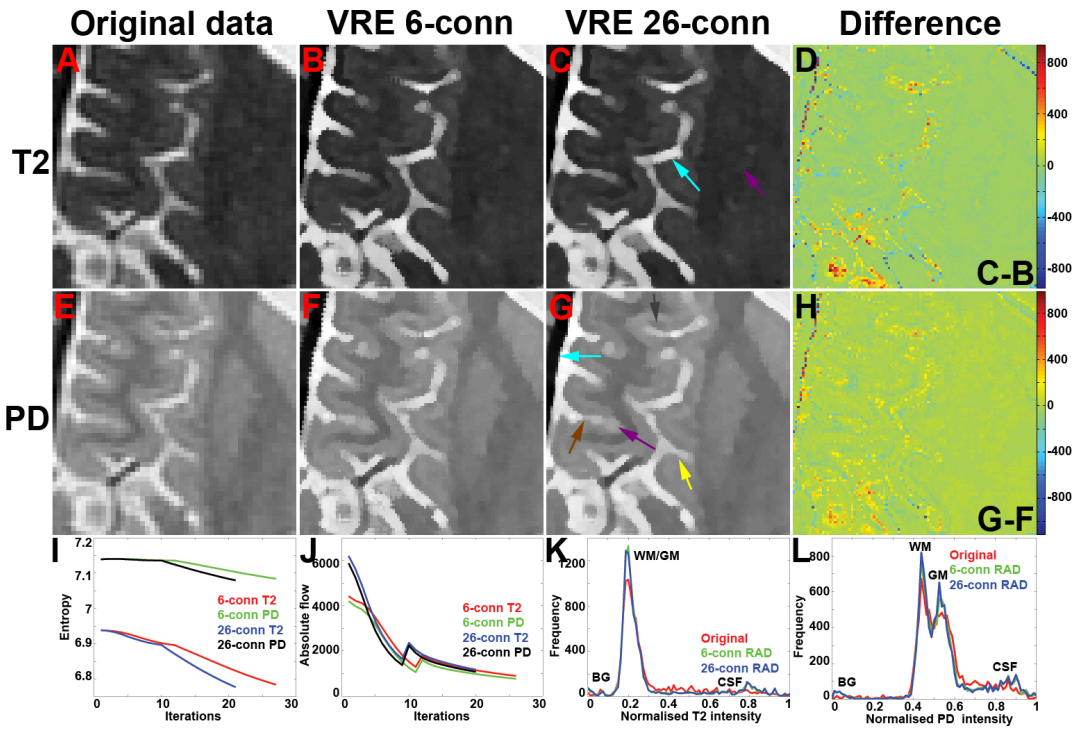

Figure 4.8: **Effects of changes to flow and gradient calculations** (LBC1936 data). **A** & **E**) T_2 and PD data for ROI 5. **B** & **F**) 6-connected, single-channel result after the reverse anisotropic diffusion phase has reached convergence. **C** & **G**) 26-connected, dual-channel result after the reverse anisotropic diffusion phase has reached convergence. **D** & **H**) Intensity differences between **B** & **C** and **F** & **G**. **I**) Total flow for each iteration for **B**, **C**, **F** and **G**. Vertical black line denotes the start of the RAD phase. **J**) Entropy for each iteration for **B**, **C**, **F** and **G**. Vertical black line denotes the start of the RAD phase. **K** & **L**) Line plots of intensity histograms for **A**, **B** and **C**, and **E**, **F** and **G**, respectively. **VRE 6-conn**: volume resolution enhancement with 6-connected, single-channel processing. **VRE 26-conn**: volume resolution enhancement with 26-connected, dual-channel processing. Coloured arrows are referenced in Section 4.3.1.

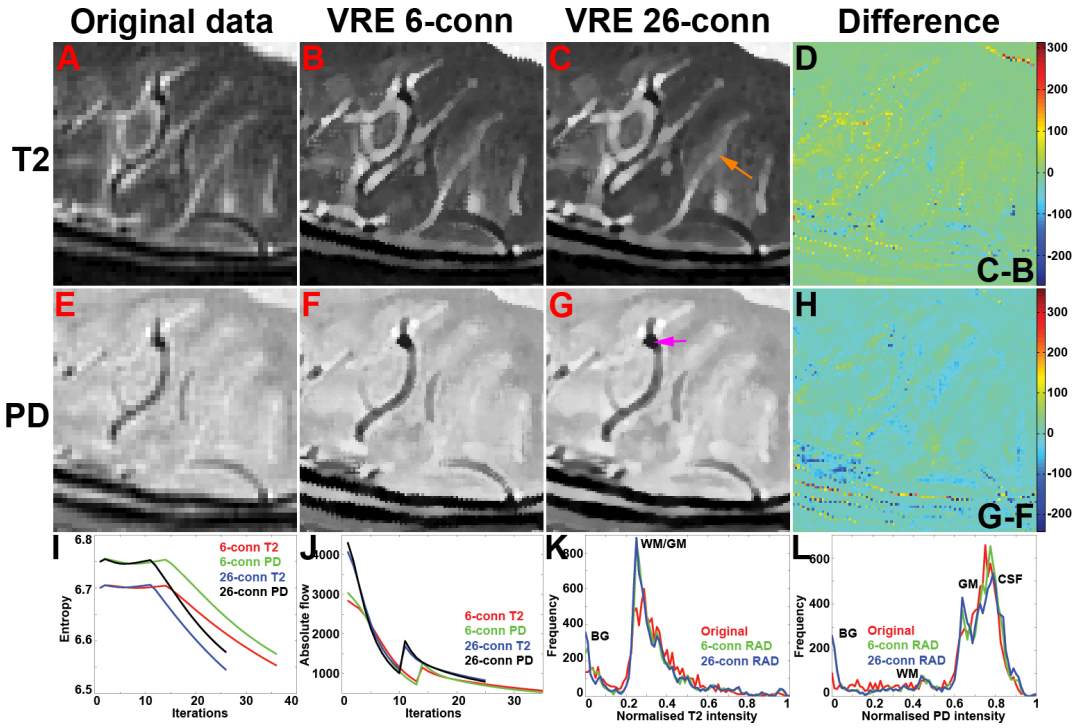

Figure 4.9: **Effects of changes to flow and gradient calculations** (young brain data). **A** & **E**) T_2 and PD data for a region from the young brain dataset. **B** & **F**) 6-connected, single-channel result after the reverse anisotropic diffusion phase has reached convergence. **C** & **G**) 26-connected, dual-channel result after the reverse anisotropic diffusion phase has reached convergence. **D** & **H**) Intensity differences between **B** & **C** and **F** & **G**. **I**) Total flow for each iteration for **B**, **C**, **F** and **G**. Vertical black line denotes the start of the RAD phase. **J**) Entropy for each iteration for **B**, **C**, **F** and **G**. Vertical black line denotes the start of the RAD phase. **K** & **L**) Line plots of intensity histograms for **A**, **B** and **C**, and **E**, **F** and **G**, respectively. **VRE 6-conn**: volume resolution enhancement with 6-connected, single-channel processing. **VRE 26-conn**: volume resolution enhancement with 26-connected, dual-channel processing. Coloured arrows are referenced in Section 4.3.1.

CSF intensities do not form for many thin cortical sulci and blood vessels (yellow arrow) due to their poor resolution. Sulci and vessels also appear wide as a result of their pure intensities not being restored; these observations show that the fidelity constraints, as local level differences (Dfn 4.1.2), are working successfully. Since intensity cannot change beyond these local levels nor be created then the putamen, thalamus and some other WM/GM boundaries in T_2 are still somewhat poorly contrasted along with some sulci in PD (purple arrows), despite contrast being readily apparent in the other image weighting factor. In general, low-magnitude noise appears to be effectively smoothed but some high-magnitude noise remains (dark gray arrow).

The graphs of changes in flow each iteration (panel J of each figure) show that the modified processing reached convergence several iterations faster for both the RD and RAD phases, and more total flow occurred during the initial iterations. Changes in entropy were much less stark (panel I), being similar during the RD phase with the final entropy only very slightly reduced in the RAD phase. There were no substantial differences in contrast between the results, borne out by the similarity of the final intensity histograms (panels K and L). Examining the changes from the original intensity histograms revealed more distinct WM and GM peaks in PD, a larger combined WM and GM peak in T_2 and the establishment of a small CSF peak in both image weighting factors for ROI 5. The young brain data exhibited similar changes to ROI 5 from its original intensity histograms but the CSF peak was much less prominent and the BG peak much more so. These differences occurred due to the presence of a large amount of GM/CSF partial volume in poorly-resolved sulci (orange arrows) that were not restored, as well as because of the presence of several fairly well-contrasted large vessels where PVE at their borders was resolved well into dark blood and light CSF sub-voxels (pink arrows).

4.3.2 Preservation and restoration of thin structures

Using a similar methodology to the previous section, modifications to the local levels and employment of shaped and oriented filtering (described in Section 4.2.3) were examined using the same ROIs, but this time without using 26-connected, dual-channel flow in order to assess the thin structure-specific processing changes in isolation. Naturally the portions of the ROIs which were not flagged in the thin structure map B (Dfn 3.4.5) – with the maps visible in Figs 3.15 and 3.17 – displayed no change in the modified processing's results (marked with an aqua arrow in Figs 4.10 and 4.11,

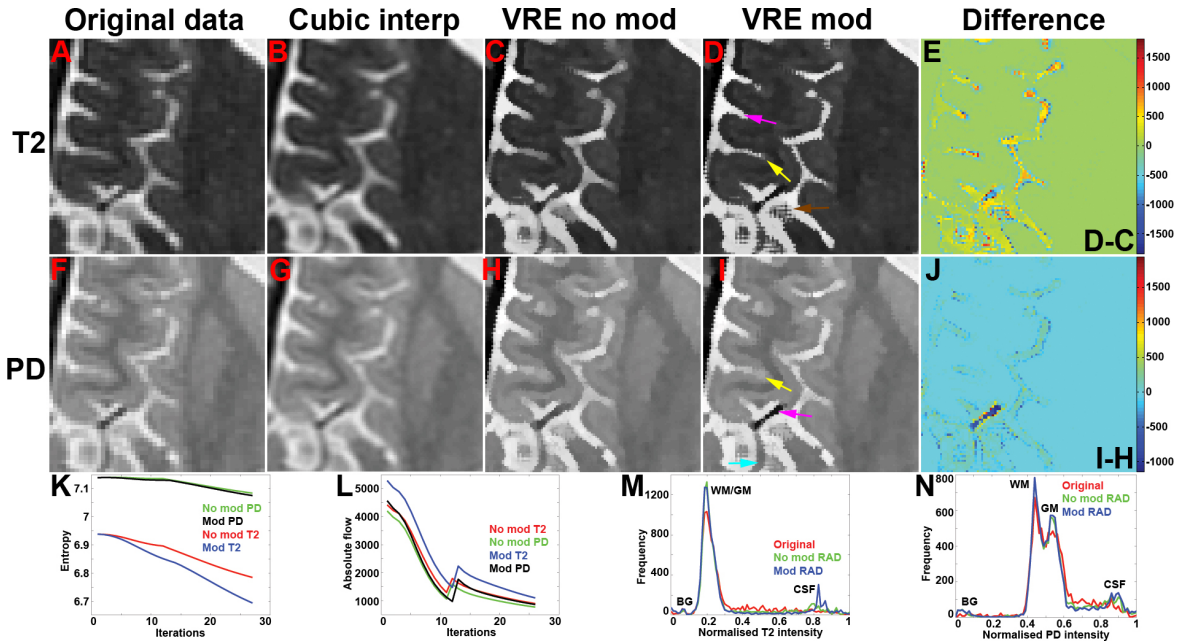

Figure 4.10: **Effect of measures to preserve and restore thin structures** (LBC1936 data). **A** & **F**) T_2- and PD-weighted data for ROI 5. **B** & **E**) Result following $2 \times 2 \times 2$ cubic interpolation. **C** & **H**) Result after the reverse anisotropic diffusion phase has reached convergence without changes to the local low and high level rank indices, local level differences or filtering in thin structure areas. **D** & **I**) Result after the reverse anisotropic diffusion phase has reached convergence using changes to the local low and high level rank indices, local level differences and shaped and oriented filtering in thin structure areas. **E** & **J**) Intensity differences between **C** & **D** and **H** & **I**. **K**) Total flow for each iteration for **C**, **D**, **H** and **I**. **L**) Entropy for each iteration for **C**, **D**, **H** and **I**. **M** & **N**) Line plots of intensity histograms for **A**, **C** and **D**, and **F**, **H** and **I**, respectively. **VRE no mod**: volume resolution enhancement without changes to the local low and high level rank indices, local level differences or filtering in thin structure areas. **VRE mod**: volume resolution enhancement with those changes. Coloured arrows are referenced in Section 4.3.2.

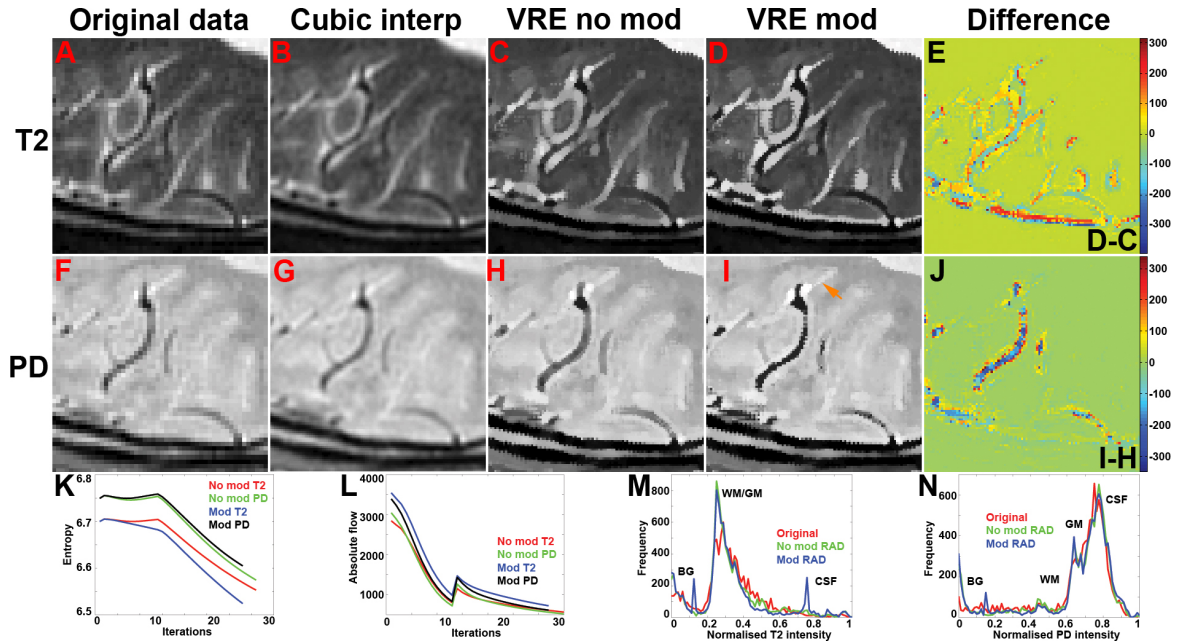

Figure 4.11: **Effect of measures to preserve and restore thin structures** (young brain data). **A** & **F**) T_2- and PD-weighted data for a region from the young brain dataset. **B** & **E**) Result following $2 \times 2 \times 2$ cubic interpolation. **C** & **H**) Result after the reverse anisotropic diffusion phase has reached convergence without changes to the local low and high level rank indices, local level differences or filtering in thin structure areas. **D** & **I**) Result after the reverse anisotropic diffusion phase has reached convergence using changes to the local low and high level rank indices, local level differences and shaped and oriented filtering in thin structure areas. **E** & **J**) Intensity differences between **C** & **D** and **H** & **I**. **K**) Total flow for each iteration for **C**, **D**, **H** and **I**. **L**) Entropy for each iteration for **C**, **D**, **H** and **I**. **M** & **N**) Line plots of intensity histograms for **A**, **C** and **D**, and **F**, **H** and **I**, respectively. **VRE no mod**: volume resolution enhancement without changes to the local low and high level rank indices, local level differences or filtering in thin structure areas. **VRE mod**: volume resolution enhancement with those changes. Coloured arrows are referenced in Section 4.3.2.

and observable in the locality of striking intensity changes in panels D and H). In the modified method's results a large proportion of sub-voxels containing sulci were correctly found to possess intensities close to the pure CSF in the ventricles and blood vessel intensities were more similar to those found in large regions of BG in the falx cerebri (not pictured) or the skull (marked with pink arrows), indicating the intensity restoration had been successful.

Not employing the changes to flow made in Section 4.3.1 meant that the consistent smoothing of structure edges between the image weighting factors, induced in Figs 4.10 and 4.11, did not occur. Some structure boundaries are inconsistently located between the image weighting factors (yellow arrows) because of the lack of dual-channel processing. Sharpening has manifested as hatching effects at some thin structure borders (brown arrow) due to much greater flow occurring only along the main data axes, which could also be remedied somewhat by using smaller values of the diffusion time-step $\Delta\tau$ (data not shown). These effects also appear striking due to the low upscaling ratio L used; at higher L there would be relatively fewer "jaggies".

Altering the contrast of the sulci and vessels should sharpen class-specific peaks (including the WM and GM classes as well, which are common neighbours to CSF and BG) in the intensity histogram and lead to a lower entropy result. Only in the young brain PD data is reduced final entropy not observed (Fig 4.11I), where some local maxima in the histogram (Fig 4.11K and L) appear to be slightly reduced. The poor contrast of sulci (particularly in PD) meant that they appeared thin when their pure intensities are restored (orange arrow). In the histograms of Fig 4.10 the CSF and BG peaks become larger (panels K and L) and the frequencies of intensities formed by partial volumes with CSF or BG are much reduced. Similar to the previous section, more flow is able to occur in the initial iterations (panel J for both figures).

4.3.3 Volume resolution enhancement accuracy

Moving from the mostly qualitative examinations of results performed in Sections 4.3.1 and 4.3.2 to quantitative analyses requires a ground truth to be established. In the case of resolution enhancement of brain volumes, the most realistic option for producing ground truth is the acquisition of the same data with a resolution L times greater, which will unfortunately have different noise characteristics as well as slightly different contrast and spatial positioning. The multi-excitation acquisition and co-registration protocols were detailed in Section 2.1 and a side-by-side comparison of

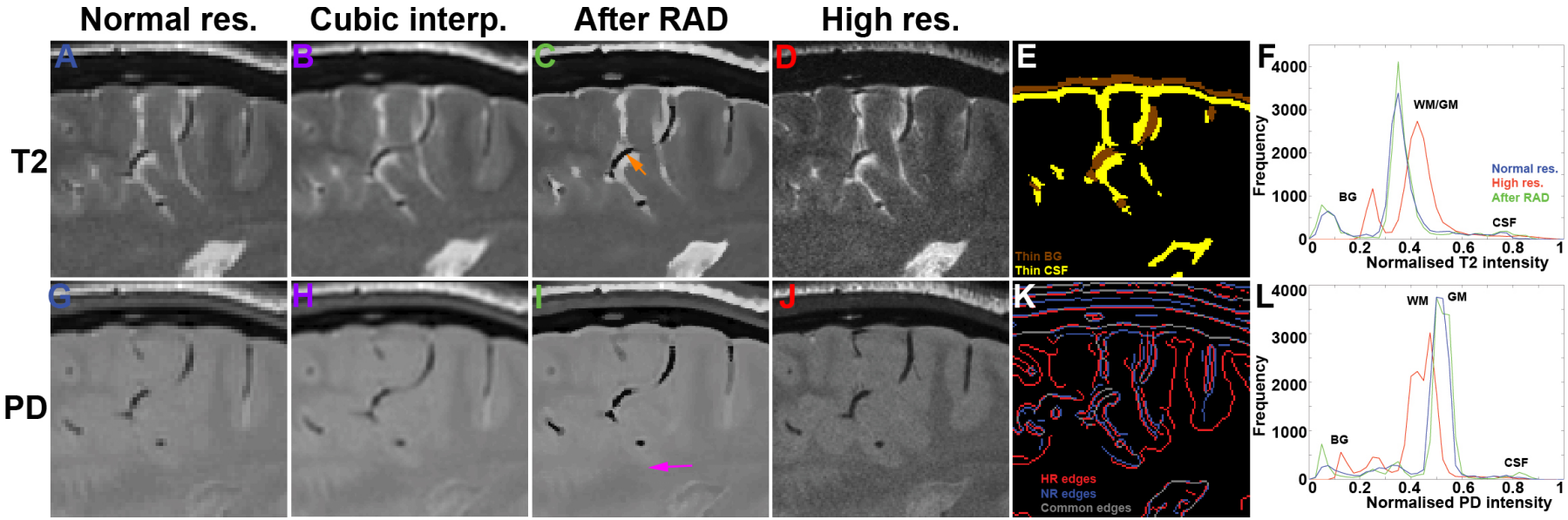

Figure 4.12: **Volume resolution enhancement of a normal resolution volume**. **A** & **G**) T_2- and PD-weighted data from the normal resolution volume ($1 \times 1 \times 2$mm voxel dimensions) in a region of temporal cortex. **B** & **F**) Result following $2 \times 2 \times 2$ cubic interpolation. **C** & **I**) Reverse anisotropic diffusion result from the normal resolution volume. **D** & **J**) T_2- and PD-weighted data from the high resolution volume ($0.5 \times 0.5 \times 1$mm voxel dimensions). **E**) Thin structure map, where brown pixels denote thin BG structures and yellow pixels denote thin CSF structures. **F** & **L**) Line plots of intensity histograms for **A**, **C** and **D**, and **G**, **I** and **J**, respectively. **K**) Edge map, where red pixels denote edges in the high resolution volume only, blue denotes those in the normal resolution volume only, and gray denotes edges common to both. **NR**, **Normal res.**: normal resolution volume. **HR**, **High res.**: high resolution volume. Coloured arrows are referenced in Section 4.3.3.

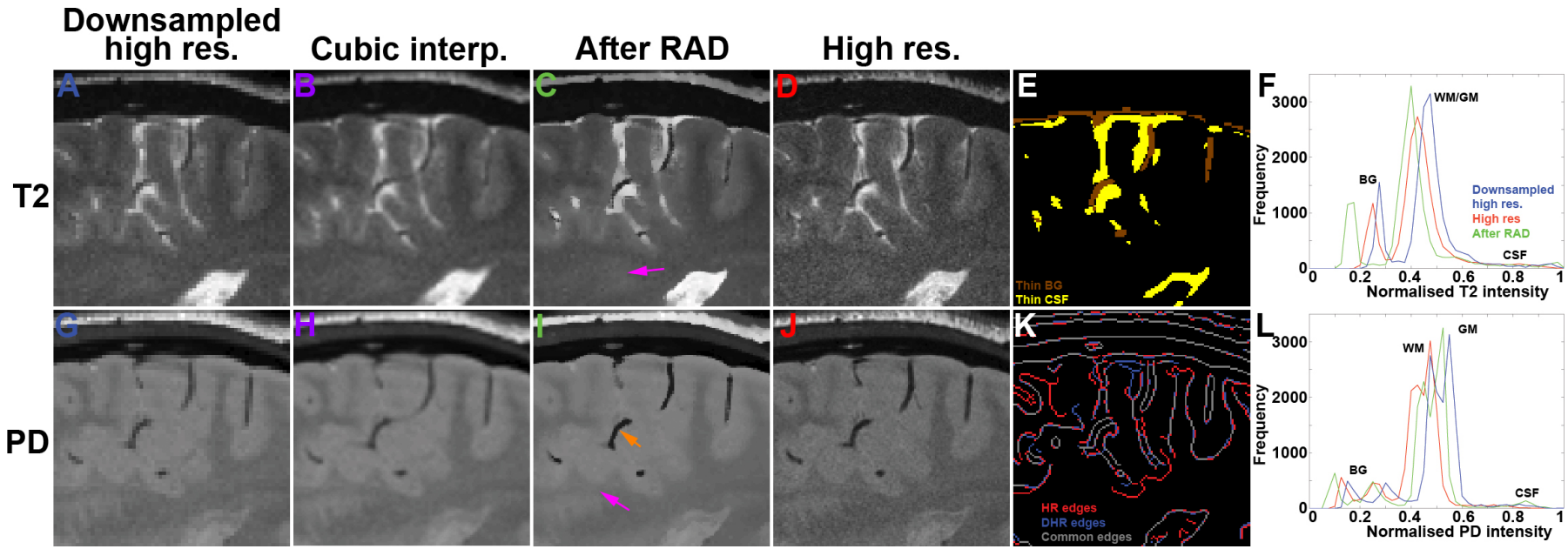

Figure 4.13: **Volume resolution enhancement of a downsampled high resolution volume**. **A** & **G**) T_2- and PD-weighted data from the downsampled high resolution volume ($1 \times 1 \times 2$mm voxel dimensions) in a region of temporal cortex. **B** & **F**) Result following $2 \times 2 \times 2$ cubic interpolation. **C** & **I**) Reverse anisotropic diffusion result from the downsampled high resolution volume. **D** & **J**) T_2- and PD-weighted data from the high resolution volume ($0.5 \times 0.5 \times 1$mm voxel dimensions). **E**) Thin structure map, where brown pixels denote thin BG structures and yellow pixels denote thin CSF structures. **F** & **L**) Line plots of intensity histograms for **A**, **C** and **D**, and **G**, **I** and **J**, respectively. **K**) Edge map, where red pixels denote edges in the high resolution volume only, blue denotes those in the normal resolution volume only, and gray denotes edges common to both. **DHR**, **Downsampled high res.**: downsampled high resolution volume. **HR**, **High res.**: high resolution volume. Coloured arrows are referenced in Section 4.3.3.

Original method of Salvado *et al.*							
ROI description	**FoM**	**T_2 MSE**	**T_2 NSDR**	**T_2 ΔE%**	**PD MSE**	**PD NSDR**	**PD ΔE%**
Normal res.	0.7507	0.0151	0.0315	2.3677	0.0126	0.0024	1.5905
D. high res.	0.8791	0.0014	0.0001	-6.5974	0.0059	0.0007	2.8237
Proposed method							
Normal res.	0.7494	0.0181	0.0459	1.7870	0.0148	0.0032	1.1321
D. high res.	0.8601	0.0016	0.0002	0.8646	0.0078	0.0023	1.7458

Table 4.1: **Error rates for volume resolution enhancement when compared with the high resolution volume**. **Normal res.** = normal resolution young brain data. **D. high res.** = downsampled high resolution young brain data. **FoM** = unified edge overlap figure of merit, where values closer to 1 are more desirable. **MSE** = mean squared error of the normalised volumes. **NSDR** = number of site differences ratio of the normalised volumes, with difference threshold 0.1. **ΔE%** = percentage reduction in entropy compared to the initial entropy value. Errors are listed to 4 decimal places. Larger percentage reductions in entropy and unified edge overlap figures of merit, and smaller mean squared errors and number of site differences ratios between the two methods' results are highlighted in red.

data acquired at both resolutions was presented in Fig 4.3. In both acquisitions GM/CSF contrast in PD was extremely low, and hence comparisons of structure edge locations were performed using the T_2 data which had at least some contrast between all the tissue types. Furthermore, since FCM clustering produced two BG clusters and a combined GM/CSF cluster because of this lack of contrast, then processing to exclude the ventricles from B was abandoned and the representative CSF intensity I^{CSF} (used in Section 4.2.3) was a little lower than would be expected. A volume similar in resolution to that used in clinical settings can also be formed by downsampling the high resolution volume by a factor of L (as outlined in Section 5.3.1); this acquisition method has the advantages that the two volumes will require no additional registration, they will not possess any contrast differences, and no non-rigid motion can occur between them.

The edge overlap figure of merit (FoM) as well as the intensity difference measures MSE and NSDR (the latter examining $> 10\%$ differences) and also the change in entropy after resolution enhancement for eleven slices surrounding the ROI pictured in Figs 4.12 and 4.13 are listed in Table 4.1 for both acquisition methods, comparing the results of the original approach [Salvado and Wilson, 2006] and those with the proposed modifications added. These metrics revealed improved edge overlap in addition to much reduced MSE and NSDR when comparing the downsampled high resolution data's results with those from the normal resolution data, which was unsurprising when considering there were no possible registration or contrast differences between the volumes in the former case. Salvado *et al.*'s original approach did produce slightly better results in all cases except for a large increase in entropy for the downsampled high resolution T_2 data, where the intensity extrema of blood vessels was smoothed to a WM-like intensity. The extrema were preserved or even restored to pure BG intensities in the proposed modified approach, and so entropy was correctly reduced. For the other metrics the original approach would naturally produce better results since it operated on single-channel data, which would lead to more accurate enhancement for the single-channel metrics. The worth of the proposed modified approach was that while it was enhancing both data channels simultaneously (with benefits to robustness to noise and structure edge localisation), its single-channel error metrics were not far behind those of the single-channel approach.

A great deal of the observations made in Sections 4.3.1 and 4.3.2 are relevant since all the proposed changes are tested together in this section. Thin structure intensities can be restored where mapped (marked with orange arrows) and their boundaries are more consistent between the image weighting factors and less jagged thanks to changes

in gradient and flow calculations. Vessels and sulci have less extreme intensities in the downsampled high resolution volume and so after resolution enhancement they generally appear thinner than for the normal resolution data's results. WM/GM boundaries remain indistinct (pink arrows) in a substantial proportion of the ROIs due to fairly low contrast between these classes.

4.3.4 Processing efficiency

Excluding the time taken for performing thin structure detection (measured in Section 3.5.5), resolution enhancement took on average 117.4 ± 5.0 seconds (mean $\pm$ standard deviation over three trials) per RD or RAD iteration when processing the volume containing ROI 5 in the LBC1936 dataset. This fulfilled the important requirement that the method process a full volume using a modern desktop PC within a few hours.

The number of iterations taken to reach convergence (defined in Section 5.3.6) were variable, but typically not much greater than 10 for both the RD and RAD phases when using the maximum possible value of the diffusion time-step $\Delta\tau$. The main factors governing the time taken per iteration were similar to those described in Section 3.5.5. In addition to the volume size, other factors contributing to determining the processing time were the upscaling ratio L, the physical dimensions of the filters used and the degree of connectedness of flow calculations, and computational complexity again appeared to be linear in these factors. At the start of processing a nearest neighbour interpolation step was performed which was guaranteed to be of linear complexity in the data size.

4.4 Discussion

The resolution enhancement methods presented in this chapter can be applied to any multi-channel, resolution-limited, untextured, volumetric data – such as structural T_2-PD MRI of the brain – where the intensities are formed by the mixel model ([Choi et al., 1991], described in Section 1.3.1). In the event that a different model of intensity formation through class mixing applies to the data, or if the imaging system PSF is found to be larger than a voxel, then a more complex model of intensity flow will need to be constrained over the size of the PSF rather than the original voxel boundaries during the RD stage (as noted in [Salvado and Wilson, 2006]) and deconvolution-based processing should be considered. The original functionality of the algorithm as a re-slicing

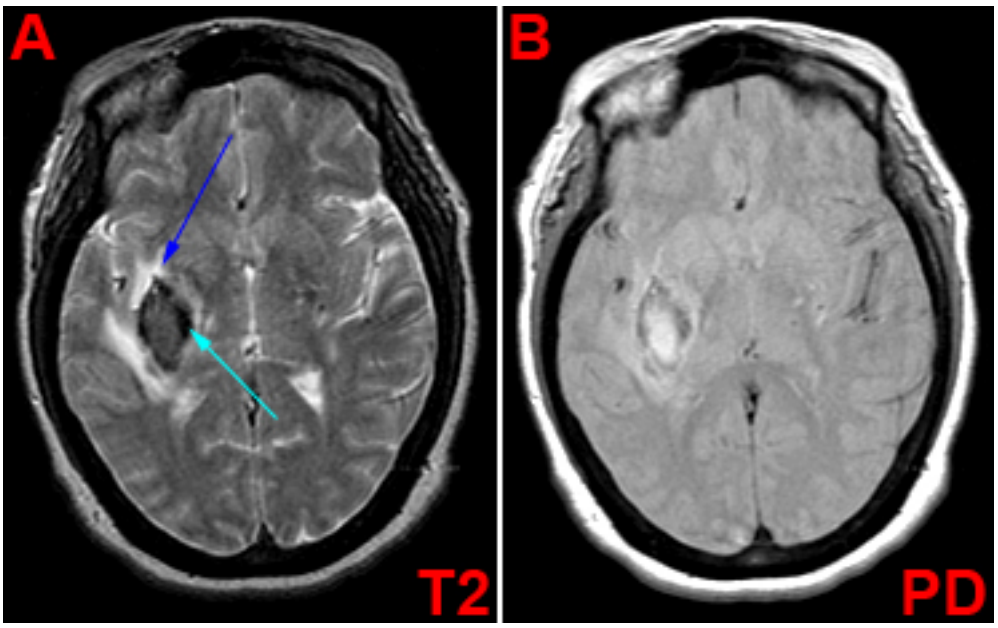

Figure 4.14: **Types of pathology problematic for volume resolution enhancement**. Diffuse oedema (blue arrow) and haemorrhage (aqua arrow) are pictured. Pictures acquired from the Whole Brain Atlas at `http://www.med.harvard.edu/AANLIB/home.html`. **A**) T_2 data. **B**) PD data.

tool for volumes with thick slices has been preserved, but volumes acquired with inter-slice gaps – effectively missing data between the slices – violate the assumption that intensities are formed by the integral of signal over the boundaries of voxels and so inpainting methods (proposed in [Bertalmio et al., 2001]) may be appropriate to fill in the missing information before diffusion takes place. Many types of pathology should be correctly processed by resolution enhancement, but non-uniformly diffuse oedemic or haemorrhagic regions (as pictured in Fig 4.14) may be incorrectly sharpened unless identified by an expert or an automatic tool (i.e. [Prastawa et al., 2004]) and masked from processing.

In evaluating the proposed methods with regard to the objectives specified in Section 4.2, concerns may be raised about altering the structure of the volume or tissue border intensities in the RAD phase (which does not employ voxel boundary constraints on flow and thus renders the original volume unrecoverable). Smoothing between sub-voxels hailing from different parent voxels is necessary to reduce noise but this operation could naturally have an impact on determining the voxels' fractional tissue class contents by examining the intensity (for example, as used directly by the segmentation method FAST [Zhang et al., 2001]). Consequently the output of the RD phase could

be chosen as the input to segmentation, where tissue boundaries and other edges have been sharpened consistently between the image weighting factors with respect to the fidelity constraints (local levels defined in Dfn 4.2.3) causing only sub-voxel changes in intensity within voxels, but where the noise has not yet been smoothed. However, it was found that the structure sizes do not substantially change after completion of the RAD phase (as evidenced by the edge overlap in Fig 4.13I and the high edge FoM values in Table 4.1) nor does the intensity (low MSE and NSDR values in this table).

Recurring themes identified in the Demonstrations (Section 4.3) include insufficient smoothing of some very noisy regions as well as poor sharpening or even some blurring of low contrast WM/GM boundaries (such as in regions of thin WM close to deep GM structures). The sensitivity to smoothing and sharpening (forward- and reverse-time diffusion, respectively) is governed by the expected noise level parameter k of the diffusion co-efficient (Eqn 4.17) but adaptation to the local structure should be implemented in the response to u, the magnitude of the intensity gradient. Appealing candidates for supplementing the gradient include wavelet-based approaches to extract feature maps of local orientation dominance, edge likelihood and phase estimates [Bharath and Ng, 2005] or to examine multi-scale correlations between wavelet co-efficients [Zhong and Sun, 2008]. Including aspects of general anisotropic [Svensson et al., 2006] or topology-preserving [Bazin and Pham, 2004] smoothing may be useful in and between larger tissue regions (such as preventing blurring of the strip of WM between the putamen and temporal cortex) and the R^B_{new} measure proposed in Dfn 3.4.3 could also be applied in this fashion. However, the act of detecting WM/GM boundaries may be so problematic that short of modifying the T_2-PD scan sequence to improve WM/GM contrast, then structural models of the anatomy (discussed in Section 6.4.1) may need to be co-registered in order to provide a spatial prior for their boundaries. T_1-weighted structural MRI data, in which the tissues are well-contrasted, could be co-registered instead with implementation of a 3-channel gradient but both methods require a perfect match of the T_1-weighted data or the atlas to the original T_2-PD data to ensure this extra information will be relevant.

Some of the poor contrast can be attributed to bias field effects – spatial variance in tissue intensities – which were not corrected during pre-processing (outlined in Section 1.1.3 and illustrated in Fig 6.1). Correcting these contrast issues would also have positive effects on later segmentation accuracy due to reducing the spread of the true pure tissue class intensity distributions and thus increasing their distinctiveness from each other. These desirable effects can also be produced by reducing noise,

which was accomplished through forward-time diffusion in low gradient regions during the RAD phase (though naturally some particularly strong noise remained). Modern MRF-based frameworks for inferring the uncorrupted data statistics (i.e. [Awate and Whitaker, 2007]) can be used in pre-processing for eliminating noise, but non-local means [Buades et al., 2005][Buades et al., 2006] presents an interesting approach by exploiting the similarity of the data in patches with independent and identically distributed noise. However, careful tuning of these methods' parameters and exclusion of regions flagged in B (Dfn 3.4.5) would be necessary to prevent the erosion of poorly-resolved thin anatomical structures.

The success of restoring pure tissue sub-voxel intensities to these structures was heavily dependent on the accuracy of their mapping in B using the methods described in Chapter 3. The coverage of these maps was found to usually slightly overestimate the thin structures, which lead to the formation of pure intensities at the core of the flagged area but only few adverse effects at the edges. Gross underestimation resulted in unsatisfactory restorative effects, as seen for blood vessels in Fig 4.13H. Poor contrast can sometimes be observed in thin structures (such as in Fig 4.11E-G) especially when thick slices have been acquired, which may advocate adaptive erosion of B at low T_{new} to the medial axis skeletons so that close-to-pure intensities form only there in order to aid visualisation and later segmentation.

The effect of thin structure restoration is the creation or heightening of BG and CSF pure class intensity peaks in the intensity histogram. In some cases the restoration contributes to the translation of the location of the peaks as normalised intensities (as observed in Figs 4.12 and 4.13) due to the creation of pure BG or CSF intensities where there were very few (or even virtually none) previously. The magnitudes of the representative intensities for pure tissue I helps to determine both the locations of these peaks and the estimation bias of the restored thin structure size: too extreme values of I will reduce the thin structures' thicknesses since more intensity will need to flow to or from the surrounding sub-voxels. Conversely, if I is not extreme enough then the original local level may not even be surpassed and the restoration will have little effect, meaning that structures may remain wide.

In the context of the literature, the proposed stand-alone resolution enhancement method differs from the unified segmentation approach [van Leemput et al., 2003] outlined in the introduction to this chapter in several ways. PVE can validly exist at the high resolution sub-voxel level allowing the proper modelling of deep GM nuclei, fidelity constraints are formed by data-driven local intensity level differences (Dfn

4.2.2) and structure maps rather than by model-driven tissue class parameters and hard neighbourhood label interactions, and resolution enhancement can be performed just once in pre-processing rather than being repeated in every segmentation iteration. The unified segmentation method of van Leemput *et al.* also requires initialisation which is not guaranteed to be robust, but since the original data is assumed to be correct – though affected by noise, bias and contrast issues – then this property propagates to the initial state of resolution enhancement (at $t = 0$) and fidelity constraints (implemented in the local level flow restrictions) ensure a degenerate result cannot be reached.

The proposed methods have been found almost as accurate (in terms of single-weighting-factor measures) with more consistent border localisation and able to restore thin structure pure tissue intensities when enhancing the resolution of low resolution or downsampled volumes (performed in Section 4.3.3), compared the original work [Salvado et al., 2006][Salvado and Wilson, 2006] on which they were based. Wider-ranging comparisons on other data would be worthwhile, as well as comparisons with related resolution-enhancement approaches. One of the key rivals would be expected to be [Fu et al., 2006], pursuing a similar strategy to Salvado *et al.* by identifying edges, textures, details and flat regions in 2D images and employing an appropriate anisotropic or isotropic shock-filter-based diffusion term in each case. However, no intensity restoration processes for poorly-resolved thin structures nor voxel-boundary-based flow constraints were considered.

The stability of the original method was maintained: flow monotonically reduced after each iteration before convergence was reached specified by the flow-based termination criterion (defined in Section 4.2.4), and the local level limits and changes to the flow calculations ensured intensity extrema did not substantially grow in size and degenerate. Another measure of improved quality – or even of convergence – can be extracted from the segmentation result of the resolution-enhanced volume since the relative proportion of pure tissue voxels (or high membership voxels in the case of fuzzy approaches) is expected to increase as PVE and noise are reduced, compared to processing the original volume. This effect is indeed observed in Fig 4.15: the tissue volumes (panel H) change at a far slower rate than the proportion of high membership elements (panel G). Entropy of the intensity histogram was found to be unsuitable for forming a termination criterion, echoing the thoughts of Salvado in [Salvado, 2006], since intensity diffusion causes gradual changes in the histogram which will not always result in sharpening the peaks every iteration.

Testing performed in Section 4.3 considered a constant value of the upscaling ratio

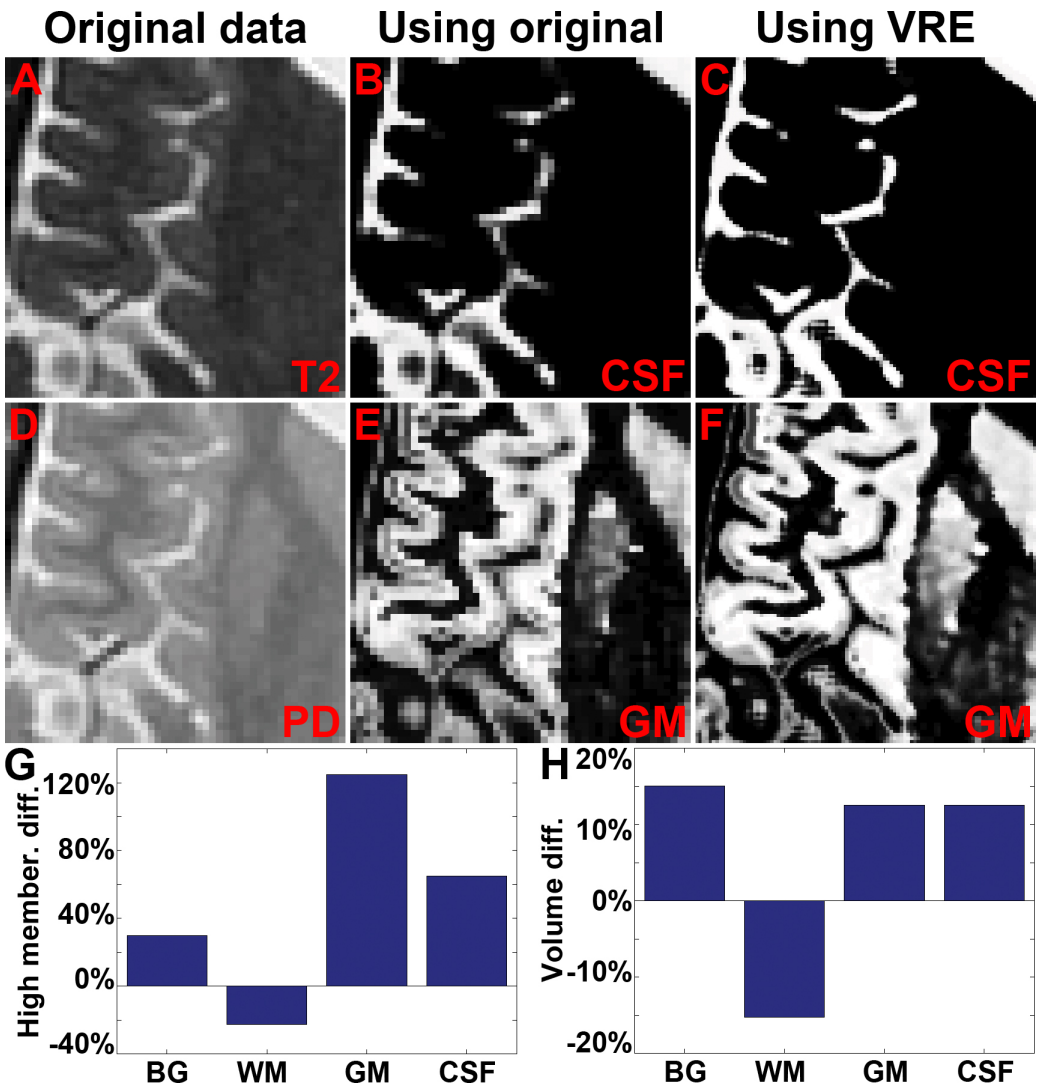

Figure 4.15: **Change in high class membership population after volume resolution enhancement**. **A** & **D**) T_2 and PD data from the young brain dataset. **B** & **E**) Class membership maps of CSF and GM created by FCM clustering (Section 5.3.2) of **A** and **E** before volume resolution enhancement. **C** & **F**) Class membership maps of CSF and GM created by FCM clustering of **A** and **E** after volume resolution enhancement. Brighter values indicate higher membership. **G**) Graph of changes in the percentage of high membership (greater than 0.9) sub-voxels in the ROI before and after volume resolution enhancement. Again, brighter values indicate higher membership. **H**) Graphs of changes in the sum of the tissue class memberships (effectively volume) in the ROI before and after volume resolution enhancement.

L, though smoother results and more impressive reduction of PVE can be produced at higher L (with higher memory and processing time requirements, as noted in Section 4.3.4). Slower diffusion with higher L may also minimise the hatching effects (visible in Fig 4.9C and G) that are stark at boundaries of contrasting structures proceeding askew to the axial plane. Better approximation of the diffusion PDE occurs at low values of the diffusion time-step $\Delta\tau$ with the trade-off of requiring more processing time, and [Zhong and Sun, 2008] noticed that $\Delta\tau$ should be smaller (roughly half of the CFL bound) to more slowly but more accurately process textured data and larger (close to the bound) for homogeneous regions. Making $\Delta\tau$ detail-adaptive, for example by examination of the local wavelet co-efficients at several scale-space levels as performed by Zhong and Sun, could be considered to prevent excessive numbers of iterations being performed with uniformly low $\Delta\tau$. Adaptation of $\Delta\tau$ could also be considered in the time domain (the iterations performed), since the most flow is expected to occur at low t when the amount of PVE in the volume is expected to be highest. Therefore $\Delta\tau$ could rise closer to the CFL bound as a function of t to ensure that flow is most accurate during the critical phase of low t, as advocated in [Gilboa et al., 2004].

Two easily-implemented steps that will increase efficiency for this more time-consuming processing involve reducing the precision of the data types used, and trimming the volume close to the parenchyma mask or ROI. The latter operation must be performed with a buffer of several voxels, or otherwise voxels near the boundary will not have neighbours that have undergone volume resolution enhancement themselves and thus cannot provide accurate local levels (Dfn 4.2.2). More substantial improvements to efficiency can be implemented with parallel processing of the gradient and flow calculations, for which each diffusion direction can largely be processed independently. Time-steps which are larger than the CFL bound – achievable in forward-time diffusion through semi-implicit additive operator splitting methods [Weickert, 1997][Barash et al., 2001] – cannot be applied to speed up reverse-time diffusion since the local level differences must still be respected in order to prevent degeneration. One strategy for increasing the speed of convergence would be increasing the exponents used in the diffusion co-efficient (Eqn 4.17) which would reduce the sensitivity to gradients but increase the amount of flow.

As more advanced MRI scanning technology becomes available, the methods presented in this chapter will not be made redundant: as noted in Section 3.6, clinicians will prioritise reductions in scan time, improvements in SNR and forming cubic voxel dimensions over improved resolution. Even at higher resolutions, volumes will still

contain a substantial amount of PVE as well as newly-visible vessels, sulci and other structures.

If slice acquisitions with a significant degree of overlap become available, then there may be some potential in incorporating elements of more common super resolution methods that rely on the input of a number of displaced acquisitions[8]. Indeed, it has been suggested that image segmentation, atlases and structural priors may contribute to regularisation of super-resolution [Greenspan, 2009] which echoes the style of approach taken in this chapter. Several algorithms have previously been applied in structural MRI, fMRI and cardiac MRI to increase resolution in the through-slice direction (i.e. in [Greenspan et al., 2002]) but since super-resolution is effectively a deconvolution of the slice profile from the data, then it may be impractical since it is sensitive to motion, the presence of noise and registration errors which can cause non-uniqueness and instability in the solution.

4.5 Summary

In this chapter two works [Salvado et al., 2006][Salvado and Wilson, 2006], combining interpolation of volumetric data with diffusion to sharpen edges while smoothing homogeneous regions at an upscaled resolution, have been extended in order to improve the quality of dual-echo T_2-PD MRI volumes used as input for the segmentation process described in Chapter 5. Since the output of the methods in this chapter display reduced noise and a greater proportion of pure tissue volume elements, the segmentation problem is simplified and a time-consuming upscaled intensity estimation step of iterations of the unified segmentation method [van Leemput et al., 2003] can be decoupled.

The modifications detailed in Section 4.2 employ the thin structure maps and adaptive filters produced in Chapter 3 so that these structures are not subject to the same sharpening and smoothing processes applied elsewhere; instead, sharpening along their thinner dimensions is subject to segmentation-based intensity limits which encourages pure tissue intensity restoration. Further changes are made to efficiently increase the number of intensity flow directions and to simultaneously process both data channels. Quantitative assessment of the method's accuracy performed on the same volume acquired at different resolutions showed more consistent edge formation than the original methods, while the errors produced remained low. In the Discussion (Section 4.4)

[8] A good review of this topic can be found in [Greenspan, 2009].

works first encountered in the literature review of Section 4.1 have been cited that may allow improvements in edge localisation, and supplementary smoothing steps to reduce noise (such non-local means [Buades et al., 2005][Buades et al., 2006]) or to preserve complicated WM/GM boundaries (i.e. [Svensson et al., 2006] and [Bazin and Pham, 2004]) have been investigated.

Chapter 5

Soft segmentation of brain tissue with thin structure preservation

In Chapter 4 several issues regarding efficiency and real-world modelling of brain tissue were identified with the soft segmentation framework presented in [van Leemput et al., 2003]: most importantly, that fractional tissue content could not be represented at the sub-voxel level and hence the sub-cortical GM (where some WM/GM boundaries cannot be observed at clinical resolutions) may be be incorrectly processed, along with poorly-resolved thin structures affected by PVE as well as tissue mixing in pathological regions. The presence of gradiated, diffuse and poorly-resolved regions which violate the piecewise-constant property, which is assumed for most brain tissue necessitate a different and practical modelling approach to be developed. The work in this chapter will also strive to achieve the goals listed in Section 1.4.

Driving the chosen approach is the desire to ensure that segmentation is deliberate as well as accurate in most tissue types and structures encountered (at different spatial scales), and also well informed. These properties have been implemented by replacing fast, greedy label optimisation in the context of neighbours (which is most typically used in the literature) with a relaxation-based method, as well as through employing the methods previously developed in this thesis:

- The effective resolution of the volumes have been increased through resolution enhancement (examined in Chapter 4). The noisiness of homogeneous regions has been reduced and the sharpness of structure edges has been improved;
- Contrasting thin structures' (found using the methods presented in Chapters 3) intensities have been preserved during volume resolution enhancement and even

restored in some cases where they were poorly resolved;

- The proportion of region boundary sub-voxels affected by PVE has been reduced during volume resolution enhancement, so more sub-voxels will possess close-to-pure intensities;

- High PPV maps have been created to provide prior knowledge of the likely locations of thin BG and CSF structures;

- Bespoke anisotropic filters (which can be appreciated as weights on information in a local neighbourhood) oriented along thin structures should encourage their correct labelling during optimisation of the labels in the context of neighbours.

An important consideration is that the segmentation method is adequately described and tested with respect to other segmentation tools, and compared against them. The performance of FAST [Zhang et al., 2001] was first investigated in Section 1.3.5 and its operation is more rigorously examined in Section 5.1 together with the work of Van Leemput *et al.* as well as other recently published structural brain MRI segmentation methods. Guidelines issued in [Udupa et al., 2006] on the quantification of the differences between segmentation algorithms in terms of accuracy, consistency, methodology and running time have been followed, and the Demonstrations (Section 5.4) show comparative testing between the proposed methods and FAST on real clinical data from the LBC1936 dataset.

This chapter is organised as follows. Since an overview of segmentation has already been presented in Section 1.3, then examinations of previous approaches to segmentation are tackled first in Section 5.1. Next, several assumptions concerning both the number of tissue classes in the parenchyma and the modelling of intensity formation are investigated. The objectives for segmentation to achieve and the details of the proposed methods are then described in Section 5.3 with notes on the motivation for design decisions. In the Demonstrations the effects of incorporating subsets of components of the complete segmentation method and its differences in performance with FAST over real MRI data are visualised and quantified. The method's performance in terms of tissue content estimation accuracy has been tested quantitatively using the BrainWeb software phantom (described in Section 1.1.6), its robustness to pathology and degeneration has been examined with quantitative testing on brain volumes with multiple sclerosis and severe atrophy, and comparisons have been made with the underlying segmentation method described in [van Leemput et al., 2003]. Finally, the

results are discussed and suggestions are made for improvements; the importance of this body of work in the context of unified segmentation approaches is discussed in the Conclusions (Chapter 6). Throughout this chapter dual-channel data variables, which contain a 2-element vector at each point, are given an overline formatting style (i.e. $\overline{X}$).

5.1 Rival brain MRI volume segmentation methods

The tools most widely used in the medical imaging community for performing soft segmentation of structural brain MRI volumes are both free to use and regularly updated: FAST[1] (previously examined in Section 1.3.5) and Segment[2] [Ashburner and Friston, 2005]. The basic processing employed by FAST is documented in [Zhang et al., 2001] but code updates have revealed changes to incorporate more accurate fractional tissue content estimation, examined later in this section. In addition, several other interesting segmentation methods will be inspected:

- The downsampling method [van Leemput et al., 2003];
- Incorporation of local gradient information [Bromiley and Thacker, 2008];
- Geodesic active contour evolution using geometric features and intensity [Huang et al., 2006];
- Fuzzy connectedness [Udupa and Saha, 2003];
- Agent-based segmentation [Scherrer et al., 2007];
- Non-parametric intensity modelling and region-based neighbourhood context [Mayer and Greenspan, 2009];
- Markov chain modelling of neighbourhood interactions during label optimisation [Bricq et al., 2008].

The performance of FAST is widely treated as a standard for comparison in the literature due to its widespread use in research environments, robustness to bias field effects, and graceful degradation of accuracy with increasing noise. It has thus been used as a benchmark for testing later in this chapter (Section 5.4). Table 5.1 lists

[1] Part of the FMRIB Software Library (FSL), available at `http://www.fmrib.ox.ac.uk/fsl/`.

[2] Part of the Statistical Parametric Mapping (SPM) package, available at `http://www.fil.ion.ucl.ac.uk/spm/`.

some published soft labelling performance statistics of segmentation tools, tested by several different authors on a standard dataset (the BrainWeb phantom [Cocosco et al., 1997], which can generate T_1-, T_2- and PD-weighted data from fractional class content maps). In these results WM labelling has generally been more accurate than GM, and where tissue content maps have been published then errors made both at the borders of convoluted cortical GM and also in labelling sub-cortical GM appropriately have given rise to this outcome (data not shown). Many methods have been optimised solely for WM and GM content quantification or discrimination, and hence investigations of accuracy over the individual CSF and BG classes have often not been performed. A variety of other hard labelling metrics were employed in these works (most commonly κ statistic scores and the Dice similarity co-efficient [Zou et al., 2004]) along with testing on other datasets (usually only accompanied by qualitative observations), but these are inappropriate for performing comparative quantitative evaluation. Therefore, discussions later in this chapter will refer mainly to their soft labelling performance on the BrainWeb dataset.

The following literature review will concentrate on contrasting the theoretical advantages associated with each method, but their common components are also worth investigating. The most popular model of intensity formation for soft labelling was the mixel model [Choi et al., 1991], which described the intensity at each voxel i as the sum of contributions from noise and tissue components; the latter is a linear combination of C tissue class means with the fractional content of each class. Each component of the label of a voxel indexed by i is $f_i = [f_{i(BG)}, f_{i(WM)}, f_{i(GM)}, f_{i(CSF)}]$ in the range $[0,1]$, and $\sum_{k \in C} f_k = 1$. The number of tissue classes which are allowed to contribute non-zero fractions to a label is known as its constrained mixel dimension. A constrained mixel dimension of 2 implies that the contributing classes with indices a and b possess fractional contents $f_{i(a)} = \alpha$ and $f_{i(b)} = (1 - \alpha)$, and for hard segmentation models α always equals 1.

The variability in intensity for pure or mixture labels were most commonly modelled using a Gaussian distribution $G(\mu, \Sigma)$ shown in Eqn 5.1 (such as in [Dugas-Phocion et al., 2004]). There are well-established, simple, robust and convergent methods for learning the parameter values Θ of Gaussian functions which goes some way to explain their common usage, but this modelling assumption is investigated further in Section 5.2.2. In Eqn 5.1 which implements the mixel model in finding the probability of label F given intensity X at the voxel indexed by i, $\overline{\mu}$ is a vector of class means and

Authors	Tool	Dataset	Error
[Bricq et al., 2008]	[Bricq et al., 2008]	T_1 3% n, 20% b	RMS WM: 0.1, GM: 0.125
[Bricq et al., 2008]	[Bricq et al., 2008]	T_1 5% n, 20% b	RMS WM: 0.12, GM: 0.15
[Bricq et al., 2008]	[Bricq et al., 2008]	T_1 9% n, 20% b	RMS WM: 0.16, GM: 0.2
[Bricq et al., 2008]	[Shattuck et al., 2001]	T_1 3% n, 20% b	RMS WM: 0.085, GM: 0.125
[Bricq et al., 2008]	[Shattuck et al., 2001]	T_1 5% n, 20% b	RMS WM: 0.125, GM: 0.15
[Bricq et al., 2008]	[Shattuck et al., 2001]	T_1 9% n, 20% b	RMS WM: 0.185, GM: 0.2
[Bricq et al., 2008]	[Bricq et al., 2008]	T_1+T_2 3% n, 20% b	RMS WM: 0.095, GM: 0.12
[Bricq et al., 2008]	[Bricq et al., 2008]	T_1+T_2 5% n, 20% b	RMS WM: 0.12, GM: 0.14
[Bricq et al., 2008]	[Bricq et al., 2008]	T_1+T_2 9% n, 20% b	RMS WM: 0.155, GM: 0.16
[Huang et al., 2006]	FAST	Average of T_1 0-9% n, 0-40% b	Absolute GM: 7.64%, WM: 17.95%
[Huang et al., 2006]	[Huang et al., 2006]	Average of T_1 0-9% n, 0-40% b	Absolute GM: 5%, WM: 5.1%
[Pham and Prince, 2004]	[Brandt et al., 1994]	T_1 3% n, no b	RMS WM: 0.126 GM: 0.139 CSF: 0.125
[Pham and Prince, 2004]	[Pham and Prince, 1999]	T_1 3% n, no b	RMS WM: 0.163 GM: 0.190 CSF: 0.200
[Pham and Prince, 2004]	[Pham and Prince, 2004]	T_1 3% n, no b	RMS WM: 0.112 GM: 0.116 CSF: 0.100
[Bazin and Pham, 2004]	[Pham, 2001]	T_1 3% n, no b	RMS All: 0.119
[Bazin and Pham, 2004]	[Bazin and Pham, 2004]	T_1 3% n, no b	RMS All: 0.126
[Bazin and Pham, 2004]	[Pham, 2001]	T_1 5% n, no b	RMS All: 0.139
[Bazin and Pham, 2004]	[Bazin and Pham, 2004]	T_1 5% n, no b	RMS All: 0.147
[Chiverton and Wells, 2008]	[Chiverton and Wells, 2008]	T_1 1% n, no b	RMS WM: 0.071 GM: 0.104 CSF: 0.061
[Chiverton and Wells, 2008]	[Chiverton and Wells, 2008]	T_1 3% n, no b	RMS WM: 0.094 GM: 0.127 CSF: 0.076
[Chiverton and Wells, 2008]	[Chiverton and Wells, 2008]	T_1 5% n, no b	RMS WM: 0.121 GM: 0.161 CSF: 0.097
[Chiverton and Wells, 2008]	[Chiverton and Wells, 2008]	T_1 9% n, no b	RMS WM: 0.158 GM: 0.215 CSF: 0.137
[Chiverton and Wells, 2008]	[Shattuck et al., 2001]	T_1 3% n, no b	RMS WM: 0.088 GM: 0.137
[Chiverton and Wells, 2008]	[Shattuck et al., 2001]	T_1 5% n, no b	RMS WM: 0.129 GM: 0.178
[Chiverton and Wells, 2008]	[Shattuck et al., 2001]	T_1 9% n, no b	RMS WM: 0.359 GM: 0.384

Table 5.1: **Rival segmentation method accuracy on BrainWeb data**. Statistics are$\pm$1 SD. **RMS**: root mean square error. **Absolute**: absolute volume error. **n**: noise level relative to maximum intensity. **b**: intensity non-uniformity level. **All**: RMS error for all tissue classes combined.

Σ is a tensor of class covariances; both parameters are summarised in Θ.

$$p(x_i|f_i,\Theta) = G(\alpha\overline{\mu}_a + (1-\alpha)\overline{\mu}_b, \alpha^2\Sigma_a + (1-\alpha)^2\Sigma_b) \tag{5.1}$$

Other types of feature distributions have also been considered. Segment [Ashburner and Friston, 2005] does not include a partial volume model and instead uses the sum of multiple normal distributions (typically 3 for GM, 2 for WM, 2 for CSF, and 5 for other tissues) to form pure class intensity distributions. In [Huang et al., 2006] as well as [Bromiley and Thacker, 2008] gradient features were incorporated to supplement the intensity data – with the latter group finding a 2% improvement in the segmentation accuracy of PD-weighted volumes – since the pure and partial volume label intensity distributions overlapped to some extent. Non-parametric densities have also been advocated for hard segmentation algorithms, since it is difficult for a single distribution shape to be valid in the presence of tissue mixing. A Parzen window kernel density estimator was used in [Mayer and Greenspan, 2009] for which the kernel design and adaptive window sizing determine the initial number of classes (maxima in the feature space) extracted via the mean-shift algorithm [Comaniciu and Meer, 2002]. These regions were then unified and clustered to produce a segmentation result with only a few classes. Non-parametric models are not so widely used for soft segmentation as it may be difficult to consistently form the intensity distribution of tissue mixtures given the pure tissue distributions.

For discovering the labelling and Gaussian class distribution parameters, the most common algorithm used is expectation-maximisation (EM) [Dempster et al., 1977]. FAST uses multiple instances of EM in sequence: initially a hard segmentation result is produced, followed by the creation of a semi-hard result where two-class mixtures are specified though not the precise fractional contents, followed by a final fractional content estimation step. This pipeline ensures that the class parameters are established before semi-hard labels are optimised in the context of neighbours, and from this homogenised result then the fractional contents of partial volume voxels can be estimated. An interesting modification of EM was presented in [van Leemput et al., 2003] where observed voxel intensities were proposed to be downsampled from a higher resolution volume, which was free from partial volume at tissue borders. This approach was notable since a hard segmentation process performed at the high resolution level can manifest as a soft one at the low resolution once downsampled. As previously noted, this attempt to circumvent the use of soft labels resulted in an inability to model structures which had been poorly resolved in the original data (resulting in partial volume

intensities existing at the chosen high resolution level) and thus caused a breakdown of their image formation model in structures such as the deep GM nuclei.

Since noise in volume intensities is a considerable confounding factor (with its magnitude sometimes surpassing the contrast between some tissue types) in the segmentation of MRI volumes, and because partial volume voxels are often at the borders between large homogeneous tissue bodies, then priors which examine the context of voxel labels in their local neighbourhood[3] are popular and effective tools for label optimisation. The Markov property states that contextual information can be entirely captured by a small neighbourhood for efficient optimisation, and it is embodied in Markov random field (MRF) theory which allows rich probabilistic models for volumetric data.

A Gibbs distribution [Geman and Geman, 1984] (Eqn 5.2) is equivalent to the joint probability distribution of the volume intensities and the labelling over its neighbourhood scheme (according to the HammersleyClifford theorem [Besag, 1974]), as a normalised negative exponential function of a prior energy function U. U can either combine single- (central-) site compatability (i.e. Eqn 5.1) between intensities (the observable field) and labels (the hidden field) with pair-site labelling compatibility (local clique potentials delivering contextual information given neighbouring voxel labels), or it can just consider pair-site labelling compatibility. The design of label matching scoring systems in the symmetric pair-site labelling compatibility function as well as the choice of scalar β[4] in the exponential – which is globally constant throughout the field in the standard Potts model – can determine the sensitivity to thin structures and noise, and even influence the proportions of tissue estimated at region boundaries in soft segmentation approaches. Z is a normalisation factor over the different labelling assignments[5].

$$P_{\beta}(F) = \frac{1}{Z} e^{(-\beta U(F))} \tag{5.2}$$

Inference of an optimal labelling configuration can be achieved in MRFs using a variety of methods, most commonly ICM (a fast but greedy algorithm) but also including belief propagation and graph cuts. Of particular importance to this thesis are non-stochastic but computationally-efficient variational methods such as mean field

[3]The relationships between a central element and its neighbours are also known as cliques, with the strength of interaction of their labelling known as potentials.

[4]Also known as a temperature parameter if β is a variable dependent on the iteration number (or time), in reference to the simulated annealing optimisation method.

[5]Z is also known as a partition function.

approaches [Zhang, 1992] to optimise the MAP criterion as well as Bethe approximation to the Gibbs free energy [Morita, 1979] to optimise the maximum posterior marginal (MPM) criterion, used by FAST.

A common approach to finding the optimum labelling is to consider the mean of Eqn 5.2 as $\langle F \rangle_\beta = \sum_f f P_\beta(f)$, where β can be considered a temperature parameter in the context of simulated annealing. Due to the interactions between the voxel labels, the calculation of the partition function involves all the possible realisations of the MRF, which makes the precise calculation of $\langle F \rangle$ not computationally feasible. The mean field theory therefore suggests an approximation to U based on the assumption that the influence of neighbouring voxels' labels f_j can be approximated by the influence of $\langle f_j \rangle$.

The mean field is usually computed by iterative procedures, which can also be implemented in parallel. As β approaches zero with increasing iterations, then the mean of the Gibbs distribution approaches its mode or the global energy minimum ($\lim \beta^{-1} \to 0 \, \langle F \rangle_\beta = F^*$). The relaxation labelling approach states that instead of minimizing U directly, the mean field $\langle F \rangle_\beta$ can be evaluated at a sufficiently high temperature and then tracked down using the continuation method: labelling ambiguity at high β is reduced to more certain label choices when $\beta \to 0$. The minimization of the MRF energy then becomes a maximization of a corresponding gain function comprising single- and pair-site compatibility functions, and $P_\beta(F)$ updates are derived from the gradient $- \bigtriangledown U(F)$. The advantages of relaxation labelling stem from the consideration of all label-label interactions in voxel neighbourhoods at every sub-iteration of the MRF optimisation (compared to greedy methods such as ICM), and in the control of the labelling ambiguity (and in turn the response to noise, fine detail and region borders) through the reduction schedule of β.

In [Chen et al., 2005] the Bethe approximation was used to approximate the Gibbs free energy rather than the mean-field approximation, and it was minimised according to its marginal posterior probability. By using the MPM then the optimum $P_\beta(F)$ was explicitly estimated on a voxel-by-voxel basis, rather than by using global optimisation methods or single voxel approximations. On simple test images this method was shown to perform better than mean field approaches as the components of a Gaussian mixture model overlapped more, and was more robust to the choice of β.

Usually the goal of modelling neighbourhood context is to reduce noise and promote homogeneity of pure tissue regions, in order to produce a robust segmentation result where spatial variation in tissue labelling is rare except at tissue borders. This

approach tends to give an impressive visual result that corresponds with models of most brain anatomy. The second phase of FAST's processing (the assignment of semi-hard labels) is typical of this approach; penalties were incurred for partial or no overlap of classes in semi-hard labels m_1 and m_2 (Eqn 5.3, where e is the inverse of the Euclidean distance to an adjacent voxel). For soft label sets a more continuous penalty can be employed, which is reduced as the geometric distance between the voxels studied increases (as they decrease in relevance to the target region) as well as when more overlap occurs between the labels' classes (as used in [Liang et al., 2007]). In the downsampling approach of [van Leemput et al., 2003] an MRF was used at the high resolution level so binary clique potentials could be used on a hard label set, but naturally this approach caused the expressiveness for any partial volume labelling at this level to be eliminated.

$$clique(m_1, m_2) = \begin{cases} 2e, & \text{if } m_1 = m_2; \\ e, & \text{if } \exists c.(c \in m_1) \wedge (c \in m_2) \\ -e, & \text{otherwise.} \end{cases} \tag{5.3}$$

Optimising MRF-based algorithms is computationally intensive and brain volumes have a large number of voxels, so time constraints lead to the selection of simplistic label optimisation methods such as the greedy iterative ICM [Besag, 1986] method used by FAST – which in turn requires strong differences to be induced in the design of its clique scoring system. To further speed up processing, in [Bricq et al., 2008] a 3D Hilbert-Peano scan was employed to transform the 3D structure of the volumes into a 1D sequence so that a Hidden Markov Chain model could be applied. However, the missing interactions between voxels in this scan – which happen since there are only two direct neighbours of each voxel in 1D – can negatively impact the segmentation results at complicated boundaries. On the other hand, more involved (and naturally more computationally-intensive) optimisation methods may improve the quality of the final result, and while fully stochastic methods such as simulated annealing and message passing may not be sufficiently time-efficient, a deterministic compromise such as relaxation labelling (implemented in [Li et al., 1997] and described in Section 5.3.4) may allow softer clique scoring systems to be designed and hence be more sensitive to small anatomical details, while still removing noise.

MRFs have also been employed in agent-based frameworks [Scherrer et al., 2007] to share knowledge of local intensity models between neighbouring agents at several spatial scales and ensure that some global consistency of the individual agents' inten-

sity models exists, which can also provide resistance to bias field effects. Some segmentation algorithms depend on deformable shape models such as geodesic active contours [Lorigo et al., 2000] or watershed transforms to provide neighbourhood context instead of MRFs, but the fronts and regions can be prone to leakage at weak or noisy edges. To combat this weakness, pre-processing of the volumes with edge-preserving anisotropic diffusion is common (i.e. [Weickert, 1998]), although difficulties become apparent in obtaining automatically-processed results which do not fundamentally alter the data.

The Segment tool does not employ shape modelling of structures nor MRF-based labelling priors, but instead requires priors describing the likely location of tissues which are acquired by registering the volume to a probabilistic tissue atlas. Segmentation quality is dependent on the relevance of the atlas to the anatomy and without a prior investigating neighbourhood context of the labelling then the quality of the segmentation can be degraded where the volume's structure deviates from the atlas (due to unique structure or pathology) or where noise occurs at region boundaries. Registration to an atlas can also be used to provide samples to calculate class distribution parameters [van Leemput et al., 1999] or to form the zero level set of an active contour [Huang et al., 2006].

Finally, the speed of the algorithms can be a concern in clinical environments, and the computation times listed for BrainWeb volumes were typically 30 minutes (i.e. [Bricq et al., 2008], [Mayer and Greenspan, 2009]) on a PC which had a modern specification at the time of publishing. In [Bricq et al., 2008] it was noted that the method presented in [van Leemput et al., 2003] (and discussed in the introduction to Chapter 4) was extremely slow due to the re-computation of the entire sub-voxel level model at each iteration. In terms of computational complexity, the most complex operations performed by FAST were the optimisation of the semi-hard labels in the context of neighbours and calculating the probabilities of each fractional tissue level at every partial volume voxel. The publically-available code revealed that the complexity was linear in the number of voxels and the neighbourhood size, which is a crucial property due to the large number of voxels in brain MRI volumes.

5.2 Investigation of data modelling assumptions

In the literature several important simplifications have been made to the brain MRI volume segmentation problem in order to minimise processing time and model com-

plexity. These include using only four tissue classes in the labelling of brain tissue, approximating the distributions of their pure intensities with Gaussian functions, and explaining the effects of tissue mixing on observed intensities using the mixel model. In the following sections, the validity of these simplifications will be tested using the expert-labelled real brain MRI dataset.

5.2.1 Validation of using four brain tissue classes to constitute brain parenchyma

The majority of parenchymal segmentation methods consider at least 3 main tissue classes: WM, GM and CSF (as outlined in Section 1.3.1). Pathological tissue classes such as tumours or lesions (discussed in Section 6.2) can also be included in models for diseased datasets, but more commonly another class (BG) comprises other tissues that do not fit the three other classes. The membership of BG can include many non-brain tissues such as the skin, eyes and so on, in which case the class intensity distribution can be modelled by uniform, non-parametric or multiple Gaussian functions. Alternatively its membership may be restricted to tissues with dark intensities which are located close to or inside the parenchyma – such as blood, the skull and air – so its intensity histogram presents a single cluster of values, like the other main tissue classes.

Assuming that skull stripping operations (described in Section 1.3.3) are effective then only parenchymal and some meningeal tissues will remain visible. To validate the four tissue class model, the anatomical locality of the *other* class (assigned to regions in the LBC1936 dataset that could not be assigned to the four main classes) was studied and 3 ROIs that show instances of *other*-labelled tissue are shown in Fig 5.1. Common to all the ROIs are glial matter and the membrane-like septum pellucidum, which consists of WM and GM laminae and sometimes a cavity, at the edges of the ventricles. In ROIs 6 and 8 the choroid plexus can be observed.

Of these structures, meningeal and membrane-like tissues appear similar in intensity to either blood, WM or GM depending on the particular layer under examination and the degree of PVE encountered. The choroid plexus that exists within the ventricles has a similar intensity to blood, and ependymocyte glia[6] at the edges of the ventricles appear similar in intensity to GM. Since only a small fraction of the brain volume is occupied by this type of glial matter, the septum pellucidum and the choroid plexus (and

[6]Neuroglial cells are almost as numerous as neurons [Azevedo et al., 2009] and include oligodendrocytes surrounding GM axons to form WM.

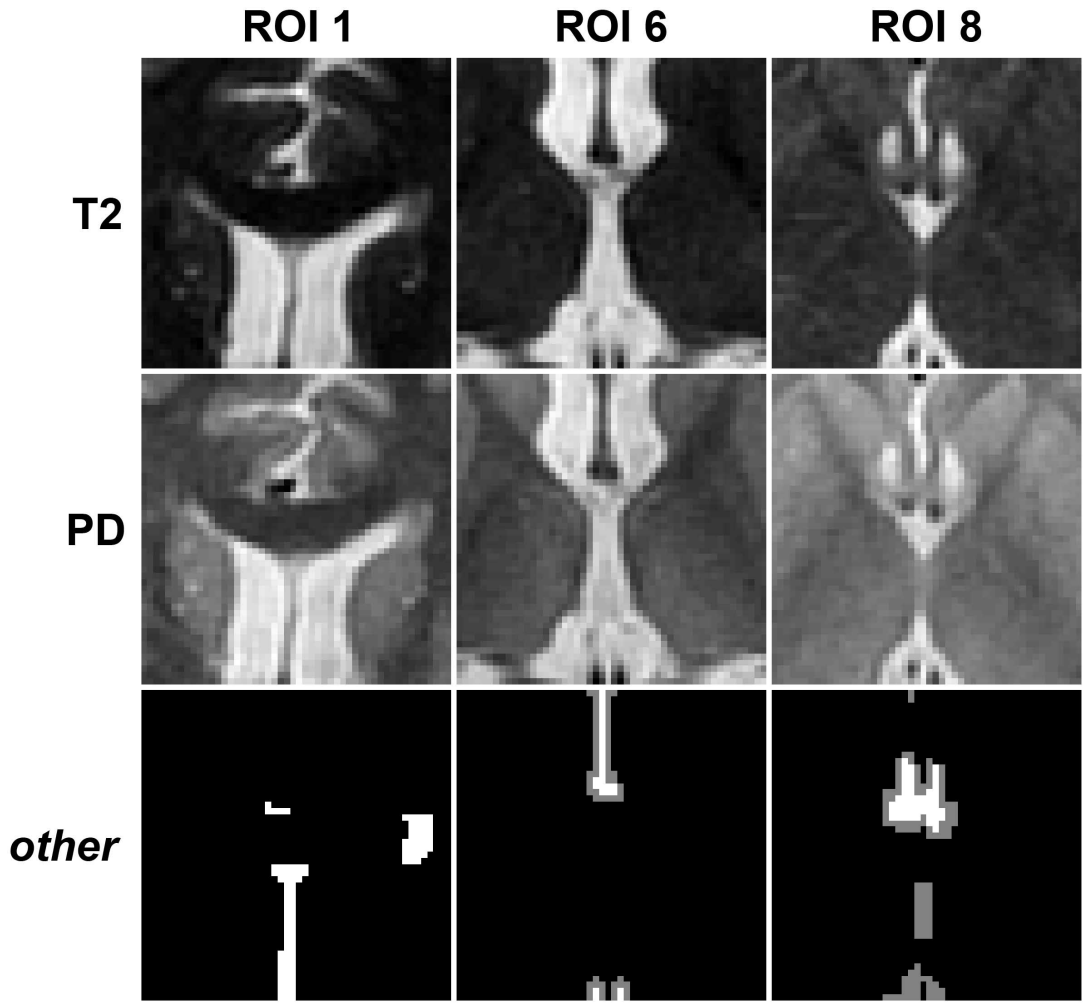

Figure 5.1: **ROIs from the LBC1936 dataset containing the $other$ class**. Fractional contents of the $other$ class are shown in the last row (white: pure, grey: partial volume, black: none).

these regions are usually not of considerable importance to clinicians) and because adding more classes to extract these regions would result in considerable overlap in intensity with the existing classes, then the four tissue class model is deemed sufficient for the task. Analyses involving *other*-labelled data points later in this chapter will therefore ignore these regions (until they can be appropriately labelled, as discussed in Section 6.4.1).

The variance of pure bias-corrected tissue intensities in MRI is largely determined by noise, sample heterogeneity and bias field effects, and so in ideal imaging conditions then sub-classes of GM are not needed despite their differences in anatomical appearance (such as the dark-coloured substantia nigra) and function (for example, hormone-producing and visual processing neurons). In WM the tracts with higher axonal density (such as the corpus callosum) appear to have darker intensities in MRI, but since blood vessels large enough to be visible at clinical resolutions typically do not pass through WM then WM/BG mixtures can be penalised during segmentation to promote proper pure WM labelling.

5.2.2 Validation of Gaussian modelling for pure tissue class intensity variation

In Section 5.1 non-Gaussian models for tissue class intensity variation were mentioned, including non-parametric distributions produced using Parzen windows and the sum of multiple Gaussian distributions. Distance metrics raised to a power have been used in clustering methods such as FCM clustering (described in Section 5.3.2), but they only indicate degrees of membership to particular clusters. In [Joshi and Brady, 2005] it has been argued that the intensity distributions (at least for hard segmentation) tend to be better modelled by non-parametric densities, though more well-established and popular processing methods exist for the mixing of Gaussian distributions in segmentation. Furthermore, non-parametric density estimation often requires manual sampling of tissue regions to ensure a representative selection is made, which is an undesirable step when aspiring to automation.

To investigate the viability of using Gaussian intensity distributions for pure tissue content, the statistics of signal acquisition can be studied. The intensities of a uniform sample acquired using MRI should be generated with a Rician distribution, or a Rayleigh distribution in regions with very low signal (such as air). For data acquired with SNR > 2 then the Rician distribution has been found to appear nearly Gaussian

[Gudbjartsson and Patz, 1995] and this figure is usually far surpassed for clinical structural brain MRI. Variability in pure tissue intensities also arises from a variety of other factors including microscopic content heterogeneity, uncorrected local magnetic field perturbations and material motion or transport, and therefore central limit theorems suggest that a Gaussian-based model may be appropriate.

In testing the viability of the Gaussian pure tissue class intensity model on the expert-labelled dataset, the data from each dual-echo T_2-PD acquisition must be considered separately since the intensity distributions are not expected to be identical between individual acquisitions. Therefore, for each volume 5000 samples (the total number of voxels in the two ROIs per volume, displayed in Section 2.1.1) were initially available. The number of valid samples is reduced by the parenchyma mask and by ensuring pure tissue label homogeneity exists in a 3×3 neighbourhood, to reduce the chance of acquiring samples which had been mislabelled by the expert at tissue borders. Pure membership of the BG class was very low and hence this class was not examined in the following analyses.

Fig 5.2 shows the intensity scatter plots for each pure tissue class label from the five volumes, and an implementation[7] of the EM algorithm [Dempster et al., 1977] was used to fit a dual-variate Gaussian density to the points in each plot. In most cases the principal axes of its plotted covariance ellipse proceeded diagonally to the top-right corner of the plot rather than radially, indicating that the off-diagonal elements of the covariance matrix are required to characterise the intensity variation (and thus just storing variances for each modality will be insufficient). The fitted distribution parameters were sometimes not located at the visually-determined data modes for GM (particularly volumes 1 and 4) and CSF (volume 3), where mislabelling of partial volume voxels as pure as well as bias field effects may have led to spreading of the intensity cluster and the creation of outliers. In general however, it can be observed that where the number of samples is sufficiently large (noted as large N in Table 5.2) a single well-defined cluster which is roughly symmetric over the principal axes of its covariance ellipse can cover model the intensity distribution well.

For more quantitative analysis of this observation then a test of multi-variate normality can be employed. As an example, an omnibus test[8] [Doornik and Hansen, 2008] for each label's intensities from individual volumes was employed to examine the in-

[7] Using code produced by Michael Boedigheimer, available at `http://www.mathworks.com/matlabcentral/fileexchange/15527`.

[8] Implemented by Antonio Trujillo-Ortiz and available at `http://www.mathworks.com/matlabcentral/fileexchange/17530`.

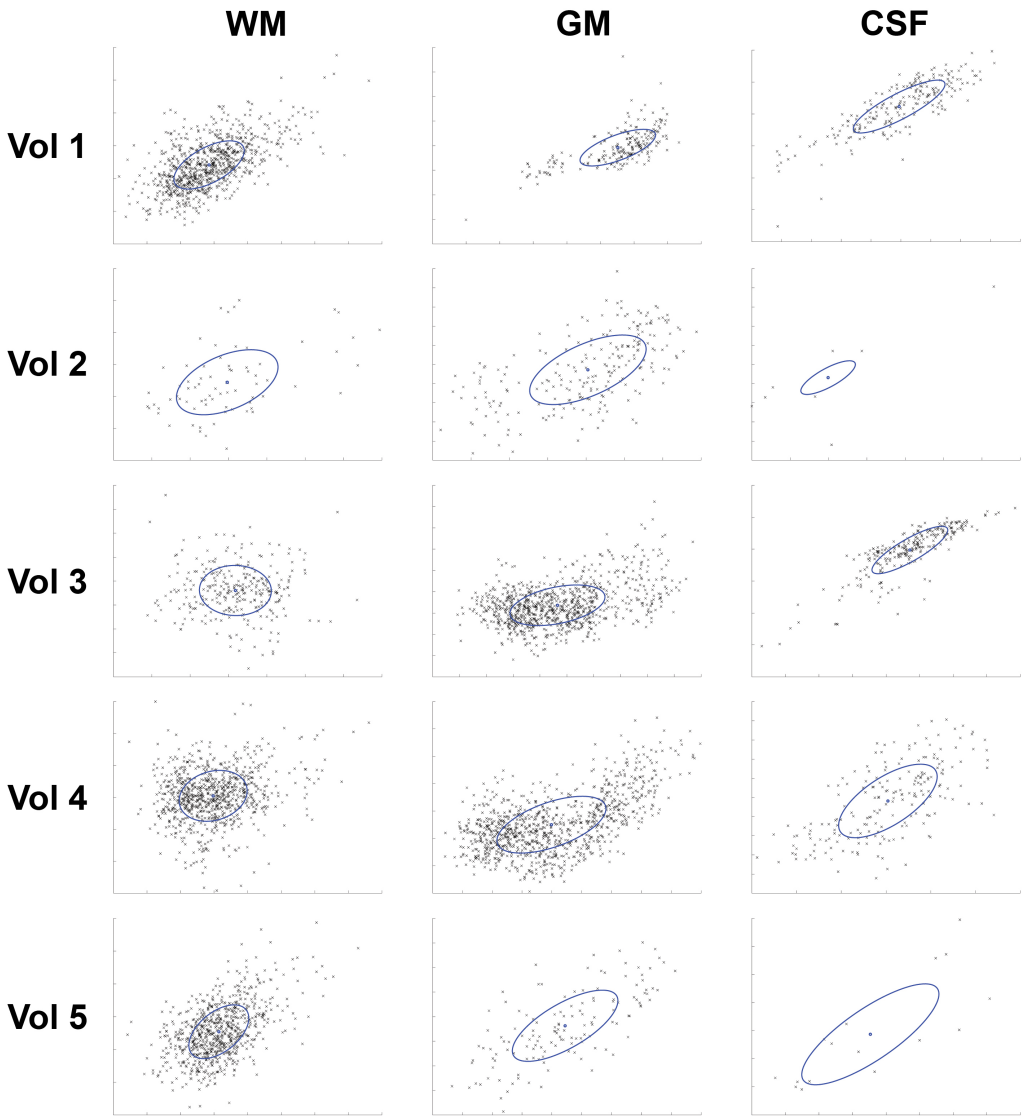

Figure 5.2: **Pure class intensity scatterplots**. Blue ellipses plot the covariance matrix of the intensities at 1 standard deviation; blue circles at their centres denote the mean. **Horizontal axes** PD values. **Vertical axes**: T_2 values. **Top to bottom**: Plots using data from pairs of ROIs in different volumes in the LBC1936 dataset. **Left to right**: Data labelled as pure WM, GM or CSF by the expert.

LBC1936	WM		GM		CSF	
volume	p	N	p	N	p	N
1	0	856	0	220	0	217
2	0.001	75	0	229	0.179	8
3	0.032	298	0	896	0	254
4	0	847	0	980	0.481	202
5	0	756	0.172	139	0.044	16

Table 5.2: **Multivariate test of normality applied to pure class intensities in ROIs of the LBC1936 dataset**. Tests were performed using the multi-variate normality omnibus test [Doornik and Hansen, 2008] and the null hypothesis of multi-variate normality is rejected at low p. p: p-value to 3 decimal places. N: number of samples.

dependence of transformed skewness and kurtosis properties. In Table 5.2 only a few of these datasets are flagged as likely to be normally distributed (large p values) despite their Gaussian-like symmetry and profile, such as for WM in volume 1. The failures can be attributed to a several factors (some reiterated from the previous paragraph):

- Low sample sizes in several cases;
- Putative mislabelled partial volume voxels and bias field effects causing scattering of the samples, and many outliers;
- Stemming from the second factor, a large modal region with similar local sample frequencies exists in many cases, indicative of poor peak definition;
- Tests of normality can often be over-sensitive, and methods such as kurtosis testing can be very sensitive to outliers [Brys et al., 2006].

These mitigating factors lead to the decision to accept the Gaussian distribution as the intensity variation model, due to the consistent qualitative observations of single (yet sometimes wide) peaks for class clusters and the apparent symmetry across the principal axes of covariance.

5.2.3 Validation of the mixel model for intensity formation in the presence of tissue mixing

Assuming that Gaussian distributions are used to model pure tissue intensity variation in the context of brain MRI as proposed in the previous section, then the mixel

model [Choi et al., 1991] can determine the likely intensities of partial volume voxels given the constituent classes and their individual fractional contents. This model was first presented in Section 5.1 but it is outlined formally in Dfn 5.2.1 for dual-channel datasets.

Definition 5.2.1 (Dual-channel mixel model) A linear combination of class mixture proportions F and the dual-channel class means $\overline{\mu}$, as well as an additional dual-channel noise vector $\overline{\varepsilon}$, produces the expected dual-channel voxel intensities $\overline{X}$.

$$\overline{X} = F.\overline{\mu} + \overline{\varepsilon} \tag{5.4}$$

The mixel model can be validated by examining the scatter plots of intensity samples from voxels given partial volume labels by the expert. These samples should lie roughly evenly along the path between the constituent pure tissue class intensity means (computed in Section 5.2.2 and visible in Fig 5.2) if the linear combination assumption holds and the distribution of fractional content is uniform. A more detailed examination of the functional relationship is not possible since the fractional contents of the tissues were not precisely estimated (as noted in Section 2.3). Only BG partial volumes were not analysed, since the BG mean was not calculated in Section 5.2.2 and also because few voxels were assigned these labels. No label homogeneity restrictions were enforced in this analysis since the majority of partial volume labels were located on the boundaries between pure tissue regions.

In Fig 5.3 it was indeed observed that the partial volume intensities lie only near the path between the pure tissue class means in most cases, and thus the mixel model is validated. This argument is less convincing for the WM/GM case with many samples extending beyond the path but due to bias field effects as well as the poor contrast between these tissues in T_2 (and particularly in posterior regions for both image weighting factors) then the potential for errors to be made in labelling as well as in estimating the mean is increased.

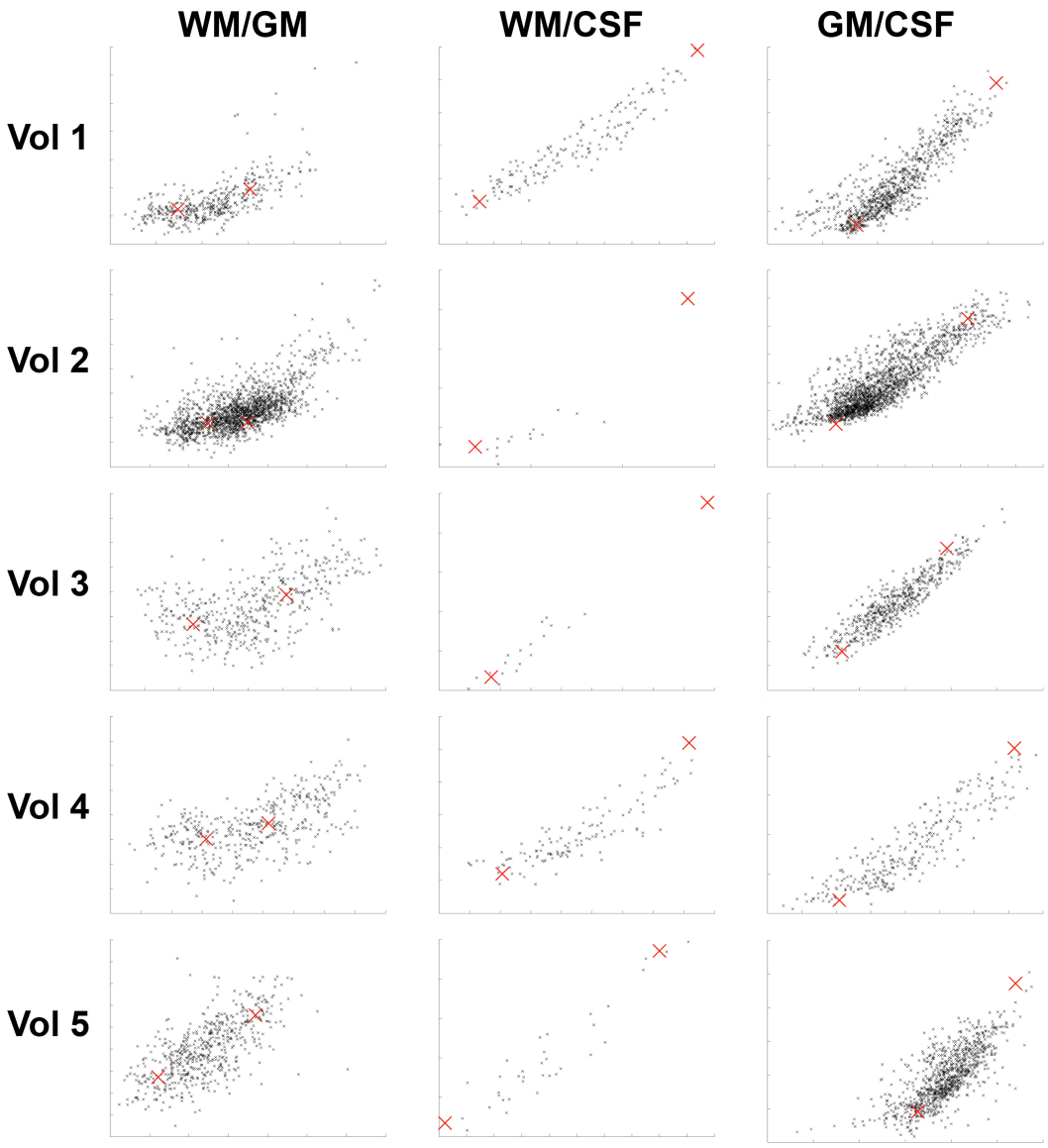

Figure 5.3: **Scatter plots of partial volume voxel intensities**. Red crosses denote putative constituent pure tissue mean intensities, taken from Fig 5.2. The leftmost cross is for the first class listed in the column header, and the rightmost cross is for the second. **Horizontal axes** PD values. **Vertical axes**: T_2 values. **Top to bottom**: Plots using data from pairs of ROIs in different volumes in the LBC1936 dataset. **Left to right**: Plots using data labelled as particular partial volumes by the expert.

5.3 Thin structure-preserving segmentation of dual-channel MRI brain volumes

Section 5.2 validated several assumptions necessary for the modelling techniques presented in this section, such as usage of the 4-class model for brain parenchymal tissue, Gaussian probability distribution models for pure tissue intensities, and the mixel model for tissue mixing inside voxels. Following initialisation (Section 5.3.2), three stages (Sections 5.3.3-5.3.5) comprise the main segmentation algorithm, forming one iteration which is repeated until convergence (defined in Section 5.3.6). The relationships between components of the entire segmentation process (including the initialisation, thin structure detection and volume resolution enhancement processing stages) are illustrated in Fig 1.23, with the basic EM framework inspired by FAST [Zhang et al., 2001]. Yet more inspiration for the chosen approach comes from improvements to the downsampling segmentation approach presented in [van Leemput et al., 2003] which are necessary to correct unresolved issues identified in both the introduction to this chapter and in Section 5.1:

- A soft label set (defined in Section 5.3.1) will permit partial volume voxels to be properly labelled at the sub-voxel level;
- Dual-channel MRI volume data should improve the distinctiveness of the four main tissue classes and increase robustness to noise;
- A relaxation-based optimisation process (detailed in Section 5.3.4) for the MRF label prior will allow more labels to be considered at each iteration compared to the greedy approach of ICM [Besag, 1986]. As a result, there may be more opportunities for label interactions, resulting in smoother labelling of homogeneous regions and improved discrimination of the constituent classes at region boundaries;
- Prior knowledge of thin BG and CSF structure locations in B (Dfn 3.4.5) may improve their discrimination in the presence of the homogenising MRF label prior (implemented in Section 5.3.3);
- Compared to the original data, resolution-enhanced data (produced using the methods presented in Chapter 4) will display sharper boundaries between tissue regions, less noise, and preserved or even restored thin CSF and BG structure intensities;

- Weighting of neighbourhoods anisotropically along thin CSF and BG structures by using shaped and oriented filters (produced as outlined in Section 5.3.4) should help to preserve their labelling during homogenising labelling optimisation processes.

5.3.1 Intensity and label models

As validated in Section 5.2.1, four tissue classes $C = \{\text{BG}, \text{WM}, \text{GM}, \text{CSF}\}$ need to be considered: white matter (WM), gray matter (GM), CSF, and a class incorporating air, the skull and blood (BG). The probability distribution of the T_2- and PD-weighted intensities of each pure – assumed 100% content – tissue class $k \in C$ is modelled with a bivariate Gaussian (settled upon in Section 5.2.2) with parameters $\overline{\theta}_k = \{\overline{\mu}_k, \Sigma_k\}$. In [Choi et al., 1991] it was reported that few voxels contain more than two tissue classes at typical clinical resolutions of brain MRI and hence a constrained mixel dimension of two is considered in the mixel model of intensity formation (investigated in Section 5.2.3): either only one class or fractions of two classes may constitute the content of a voxel.

Labels $\hat{F}$ at the original volume resolution are formed by downsampling the sub-voxel level labels F, and hence the granularity of the voxel label set $\hat{M}$ is much finer than that of M for the sub-voxel level. Given that the dual-channel resolution-enhanced volume has been produced (as described in Chapter 4) through upscaling by a ratio L to create N sub-voxels, there are $l = \prod_{d=1}^{3} L_d$ child sub-voxels for each parent voxel. $\hat{N} = \frac{N}{d}$ voxel labels can be re-formed by downsampling the sub-voxel level labels (Eqn 5.5); a similar method can be used to produce a volume $\overline{Y}$ at the original resolution by downsampling the resolution-enhanced intensities of $\overline{X}$.

$$\hat{F}_i = \frac{\sum_{j \in child(i)} \overline{F}_j}{d} \tag{5.5}$$

The cardinality of M is governed by M^{frac}, the set of available fractional levels that each tissue class can assume. The minimal set $\{0,1\}$ for M^{frac} forms a hard segmentation model at the sub-voxel level, similar to [van Leemput et al., 2003], while adding other fractions to M^{frac} allows soft segmentation to be performed at this level. Naturally, both models become soft at the voxel level once the sub-voxel labels are downsampled.

5.3.2 Initialisation of class parameters

A robust initial estimate of the pure tissue class distributions will minimise the number of iterations required for convergence and also increase the chance of finding a good optimum of the segmentation energy function, which is important for iterative estimators such as EM [Dempster et al., 1977]. Either registered tissue prior maps or variants of k-means clustering have been used in most popular segmentation algorithms (i.e. [Zhang et al., 2001], [Huang et al., 2006], [Bricq et al., 2008]) to initialise the mean and covariance parameters in $\overline{\theta}$. Fuzzy c-means (FCM) clustering, described in [Bezdek and Pal, 1991], can estimate normalised class memberships U by the distance of features from class centroids V. Its chief advantages lie in its fuzzy output U from which likely pure (high membership) sub-voxels can be selected, as well as the simplicity and high efficiency of optimising its objective function J using Lagrange multipliers.

The implementation of FCM in [Ahmed et al., 2002] (with J adapted for dual-channel data shown in Eqn 5.6) incorporates influence on the distance function from the neighbouring (Nh) sub-voxels' intensities in order to improve robustness to noise as well as better discriminate mixture components at region boundaries. In noisy images the strength of the neighbourhood influence α can be increased to produce a more homogeneous result, at the expense of further eroding thin structures. p controls the allowed fuzziness of the computed memberships and $p = 2$ is held to be the most common choice for volumes with moderate noise[9]. Convergence of FCM was assessed by thresholding the changes to $\overline{V}$ between iterations as a very low percentage of the maximum intensity of each volume. Pre-defined fractions $\overline{V}^{init}$ of the 99^{th} percentile ranked intensity for each volume initialised $\overline{V}$, which adaptively helped to improve robustness to large outliers in the intensities.

$$J = \sum_{i=1}^{N} \sum_{k \in C} u_{ik}^{p} \left\| x_i - v_k \right\|^2 + \frac{\alpha}{size(Nh)} \sum_{i=1}^{N} \sum_{k \in C} u_{ik}^{p} \sum_{j \in Nh(i)} \left\| x_j - v_k \right\|^2 \tag{5.6}$$

Since BG is poorly represented within the brain parenchyma compared to the other classes then cubic regions of voxels containing air are sampled from the corners of the volume, which are assumed to have pure BG class membership and thus can represent parenchymal BG. To form an estimate of the other class parameters $\overline{\theta}$ for the initial E-step (described in Section 5.3.3), U was thresholded by $0 \ll \pi^{FCM} \leq 1$ to produce a map of likely pure tissue voxels, which was then morphologically eroded to select

[9] $0 < p < 2$ is appropriate for low-noise conditions, not commonly encountered in MRI.

samples further toward the core of large tissue regions that were even more likely to contain pure tissue. The maximisation step of the EM algorithm (designed in Section 5.3.5) was then employed to learn the initial $\overline{\theta}$ from these samples, replacing the conditional expectation of class-specific contributions to the intensity at each voxel $\overline{S}$ with the intensities from the samples collected and the fractional labels $\hat{F}$ with a vector containing unity throughout.

5.3.3 Expectation step

The goal of the expectation step (abbreviated to E-step henceforth) for each iteration of the EM algorithm, indexed by t, is to prescribe the most likely label assignment F^t to all sub-voxels given the current estimate of the parameters $\overline{\theta}^{t-1}$. This is achieved by forming the conditional expectation of the complete data log likelihood function Q (Eqn 5.7) over F based on $\overline{\theta}^{t-1}$ and the observed dual-channel resolution-enhanced intensities $\overline{X}$ (part of the complete data pair $(\overline{X}, F^t)$). The thin structure map B (Dfn 3.4.5), the constrained mixel dimension *cmd*, and the shaped and oriented filters ϕ (prescribed in Section 3.4.4) also contribute to the likelihood function. Using Bayes' rule and the law of total probability then the conditional probability of the labelling $p(F^t|\overline{X}, \overline{\theta}^{t-1}, B, \phi, cmd)$ (Eqn 5.8) used in Eqn 5.7 can be expressed solely in terms of the conditional probability of the intensities $p(\overline{X}|F^t, \overline{\theta}^{t-1})$ and the prior probability of the labelling $p(F^t|B, \phi, cmd)$.

$$\begin{aligned} Q(\overline{\theta}|\overline{\theta}^{t-1}) &= E_F(\log\left[p(\overline{X}, F^t|\overline{\theta}, B, \phi, cmd)\right]|\theta^{t-1}, B, \phi, cmd) \\ &= \sum_F p(F^t|\overline{X}, \overline{\theta}^{t-1}, B, \phi, cmd)\log\left[p(\overline{X}|F^t, \overline{\theta})p(F^t|B, \phi, cmd)\right] \end{aligned} \quad (5.7)$$

$$p(F^t|\overline{X}, \overline{\theta}^{t-1}, B, \phi, cmd) = \frac{p(\overline{X}, F^t|\overline{\theta}^{t-1}, B, \phi, cmd)}{p(\overline{X}|\overline{\theta}^{t-1}, B, \phi, cmd)} = \frac{p(\overline{X}|F^t, \theta^{t-1})p(F^t|B, \phi, cmd)}{\sum_{F'} p(\overline{X}|F', \overline{\theta}^{t-1})p(F'|B, \phi, cmd)} \quad (5.8)$$

The conditional probability of the intensities given label m (Eqn 5.12) is obtained using a bivariate Gaussian distribution G (Eqn 5.11), normalising the result for each m over the sum of results using the entire label set M. The formation of label-specific means and covariances in $\overline{\theta}$ (Eqns 5.9 and 5.10) follow the mixel model of intensity

formation as described in Section 1.3.

$$\overline{\mu}_m = \sum_{k \in C} m_k \overline{\mu}_k \tag{5.9}$$

$$\Sigma_m = \sum_{k \in C} m_k \Sigma_k \tag{5.10}$$

$$G(\overline{x}, m, \overline{\theta}) = \frac{1}{2\pi} |\Sigma_m|^{-\frac{1}{2}} e^{-\frac{1}{2}(\overline{x}_i - \overline{\mu}_m)^T (\Sigma_m)^{-1} (\overline{x}_i - \overline{\mu}_m)} \tag{5.11}$$

$$p(\overline{x}_i | f_i^t, \overline{\theta}^{t-1}) = \frac{G(\overline{x}_i, f_i^t, \overline{\theta}^{t-1})}{\sum_{m \in M} G(\overline{x}_i, m, \overline{\theta}^{t-1})} \tag{5.12}$$

The prior probability of the labelling is independent of $\overline{\theta}$ and since B, ϕ and cmd are not related to each other then $p(F|B, \phi, cmd) = p(F|B)p(F|\phi, cmd)$. $p(F|B)$ is a thin structure localisation prior (Dfn 5.3.1) exploiting the high PPV of B, and $p(F|\phi, cmd)$ is a neighbourhood label context prior (tackled in Section 5.3.4).

Definition 5.3.1 (Thin structure localisation prior) The **thin structure localisation prior** $p(f_i|b_i)$ initially assumes that all labels in M are equally likely in the sub-voxel indexed by i, but B can specifically flag non-zero fractions of BG and CSF. Labels with no content of the respective tissue class in regions flagged by B are therefore penalised by a constant scalar multiplier $0 \leq pen < 1$.

$$p(f_i|b_i) = \begin{cases} pen, & \text{if } \exists k \in \{BG, CSF\} \text{ then } (f_{ik} = 0) \wedge b_i^k; \\ 1, & \text{otherwise.} \end{cases} \tag{5.13}$$

5.3.4 Relaxation step

The E-step (described in Section 5.3.3) contains the relaxation step (abbreviated to R-step henceforth) to optimise the energy of the MRF-based neighbourhood label context prior $p(F|\phi)$. $p(F|\phi)$ aims to promote spatial coherence of the labelling in the presence of noise as well as to help inform which are the likely constituent classes at region borders, and in many other works MRFs optimised greedily by ICM [Besag, 1986]

have presented a convenient method for modelling such 3D context. MRFs can be characterised using a Gibbs distribution [Besag, 1974] with an energy function U (as part of Eqn 5.14) which contains a recursive element since an iterative optimisation process (with sub-iterations indexed by z) is desired.

$$p(f_i^z|\phi_i, p(F^{z-1})) = \frac{e^{-\beta_z U(f_i^z, b_i, \phi, p(F^{z-1}), cmd)}}{\sum_{m \in M} e^{-\beta_z U(m, b_i, \phi, p(F^{z-1}), cmd)}} \tag{5.14}$$

In contrast to ICM, relaxation labelling [Rosenfeld et al., 1976] allows all possible label interactions in cliques to be considered, rather than assuming the most likely label has been assigned to each sub-voxel in the previous sub-iteration. Smaller values of the sub-iteration-dependent function β_z in Eqn 5.14 will cause the exponential function to produce more extreme values, rendering the dominance of particular label probabilities more ambiguous at higher β_z. This situation is beneficial at the start of optimisation – at low z – since intensity- (Eqn 5.12) and tissue prior-based (Dfn 5.3.1) label probabilities can be somewhat equalised in magnitude to allow a wider range of likely labels to be considered, and hence improve the potential for correct labelling in regions with low intensity contrast and areas with high spatial variation in tissue content. The labelling can then be refined to single candidates with a probability close to 1, normalised across the label set at each sub-voxel, at high z and low β_z once a number of chances for the labels to change has occurred.

The form and optimisation of U (Eqn 5.15) followed the mean field implementation detailed in [Li et al., 1997], as the derivative of a gain function balancing unary (central site r_i) and binary (clique pair site $r_{(i,i')}$) compatibility functions. Both r functions returned non-negative and zero-maximum results in order to maintain feasibility while performing gradient descent. In light of the prior information available on thin structure presence and shape, as well as the need to constrain the mixel dimension at the voxel level, these functions needed to be redesigned.

$$U(m, b_i, \phi, p(f_i^{z-1}), cmd) = r_i(m, b_i) + 2 \sum_{i' \in Nh(i)} \sum_{m' \in M} r_{(i,i')}(m, m', \phi_i, cmd) p_{z-1}(m'_{i'}) \tag{5.15}$$

The unary r (Eqn 5.16) was relevant to only the central sub-voxel of the neighbourhood and so it was based upon the conditional probability of the labelling (Eqn 5.8) with the dependence on neighbourhood context removed. Since both the thin structure localisation prior $p(F|B)$ (Dfn 5.3.1) and the conditional probability of the intensities

$p(\overline{X}|F,\overline{\theta})$ (Eqn 5.12) remained constant during the R-step, in turn the unary r was also constant over all z and it only needed to be calculated once per E-step.

$$r_i(m,b_i) = 1 - [p(\overline{x}_i|m,\overline{\theta}^{t-1})p(m|b_i)] \tag{5.16}$$

The binary r (Eqn 5.17) scored the interactions between each label of the central sub-voxel i and its neighbours i' (clique potentials), weighted by the value in the relative location of i' from i in the shaped and oriented filter of i in the filter set ϕ (or isotropic filter for sub-voxels not flagged in B). Anisotropic weighting has been employed to help preserve the labelling of thin structures during neighbourhood optimisation, by giving greater weight to nearby sub-voxels which are putatively part of the same thin structure. Two other factors contributing to the binary r are the clique potential ζ and the mixel dimension constraint ϑ.

$$r_{(i,i')}(m,m',\phi_i,cmd) = \max(0,\phi_i(i')\vartheta_{(i,i')}(m,m',cmd)\zeta(m,m')) \tag{5.17}$$

The form of the clique potential scoring function ζ (Eqn 5.18) was changed from the Euclidean distance between labels, suggested in [Li et al., 1997], since it did not produce large enough differences between different labels and led to poor label homogeneity in uniform regions. Instead, the more extreme function used by FAST (Eqn 5.3) was adapted for continuous mixtures by changing the final partial tissue content matching case to be based upon the normalised absolute difference. Extreme responses to perfectly-matching labels (case 1) or labels with no class overlap (case 2) were preserved to reward label homogeneity in uniform regions and to help disambiguate the correct class constituents at structure boundaries. However, the inclusion of these cases means that allowed soft tissue fractions in M^{frac} should not be very close to pure nor absent tissue content levels (1 or 0)[10] so the first two cases can be matched more often.

$$\zeta(m,m') = \begin{cases} -1, & \text{if } m = m'; \\ 2, & \text{if } (m \neq m') \wedge (\forall k \in C, [m_k > 0] \rightarrow [m'_k = 0]); \\ \frac{\sum_{k\in C} m - m'}{2}, & \text{otherwise.} \end{cases} \tag{5.18}$$

The mixel dimension constraint ϑ preferred that only pure tissue or two-class mixtures should exist within any voxel (after effectively downsampling its child sub-voxels' labels) to help this modelling assumption. ϑ returned an extreme penalty

[10] For example, $M^{frac} = \{0, 0.05, 0.5, 0.95, 1\}$ would be unacceptable since 0.05 and 0.95 are too close to 1 and 0.

($cmdpen > 1$ when $\zeta(m,m') \geq 0$, or $\frac{1}{cmdpen}$ when $\zeta(m,m') < 0$) when two conditions held: the path between sub-voxels i and i' did not cross a voxel boundary at the original resolution – i and i' were both child sub-voxels of the same parent voxel – and the union of labels m for i and m' for i' resulted in more than the number of non-zero class contributions specified by the constrained mixel dimension. ϑ is the same within every voxel and remains constant during the entire EM algorithm, and so it can be pre-computed.

$$\vartheta_{(i,i')}(m,m',cmd) = \begin{cases} cmdpen, & \text{if } (\text{parent}(i') \neq \text{parent}(i)) \wedge \\ & (\sum_{k \in C} [m_k + m'_k > 0] > cmd) \wedge (\zeta(m,m') \geq 0); \\ \frac{1}{cmdpen}, & \text{if } (\text{parent}(i') \neq \text{parent}(i)) \wedge \\ & (\sum_{k \in C} [m_k + m'_k > 0] > cmd) \wedge (\zeta(m,m') < 0); \\ 1, & \text{otherwise.} \end{cases} \tag{5.19}$$

Following convergence of the R-step (with a probability saturation criterion described in Section 5.3.6) a post-processing step was performed on the labelling to enhance the pure tissue content. Since diffuse structures are not expected within the brain parenchyma, within any large body labelled with partial volume labels should exist a pure tissue core. For example, BG/CSF labels are typically assigned to blood vessels due to their small radial size, and so any sub-voxel which is entirely surrounded by partial volumes of BG can be assigned a pure BG label since only blood can exist at the centre of the vessel.

5.3.5 Maximisation step

Once the E-step is complete through convergence of the R-step, the maximisation step (abbreviated to M-step henceforth) maximises the expected value of the log likelihood function $Q(\overline{\theta}|\overline{\theta}^t)$ (Eqn 5.7) with a new estimate of the class parameters $\overline{\theta}^{t+1}$. The M-step derived for a soft segmentation model in a series of works [Liang et al., 2003][Liang et al., 2007][Liang and Wang, 2009] was selected since volumetric estimates of class contents have been produced in the E-step (Section 5.3.3). The derivations for dual-channel data remain almost the same, and so the bulk of the intermediate steps have not been reproduced here. Before the M-step begins, the sub-voxel intensities and labels are downsampled to the original resolution (using Eqn 5.5) causing the

resulting voxel label set $\hat{M}$ to be more fine-grained than M. Hence, this allows more precise re-estimation of $\overline{\theta}$ and it also reduces the computation time for the M-step.

The computation of the update rule is performed by differentiation of the log likelihood function (Eqn 5.7) with respect to the individual parameters and setting the gradient equal to 0. The conditional expectation of class-specific contributions to the intensity at each voxel $\overline{S}$ are derived from the voxel intensities $\overline{Y}$ (Eqn 5.20) in an identical manner to that outlined by Liang *et al.*, and hence the derivation is not reproduced here. Using $\overline{S}$ instead of $\overline{Y}$ allows the update rules for $\overline{\mu}$ and Σ (Eqns 5.21 and 5.22) to appear very similar to those used in hard segmentation. As noted in Section 5.3.2, the BG class is not represented well in the brain parenchyma and will not even exist in some sub-volumes, and so $\overline{\theta}_{BG}$ remained constant from the initialisation stage.

$$\overline{Y} = \sum_{k \in C} \overline{S}_k \tag{5.20}$$

$$\overline{\mu}_k^{t+1} = \frac{\sum_{i \in N^{valid}} \overline{s}_{ik}^t}{\sum_{i \in N^{valid}} f_{ik}^t} \tag{5.21}$$

$$\Sigma_k^{t+1} = \frac{1}{\#N^{valid}} \sum_{i \in N^{valid}} \frac{(\overline{s}_{ik}^t)^2 - 2\overline{s}_{ik}^t f_{ik}^t \overline{\mu}_k^t + (f_{ik}^t \overline{\mu}_k^t)^2}{f_{ik}^t} \tag{5.22}$$

There is greater potential for labelling errors to occur at region boundaries than elsewhere, since there is high spatial variation in the labelling which may be affected more by noise. To prevent these voxels from contributing to the class parameter re-estimation process then the labelling was checked for coherency in a local neighbourhood ω around each voxel indexed by i ($\forall i' \in \omega(i), \hat{f}_i = \hat{f}_{i'}$) forming N^{valid}, a subset of N.

5.3.6 Convergence criteria

Both the E-step and R-step are iterative procedures for which automatic processing demands formal convergence criteria to be specified. In [Li et al., 1997] it was suggested that a very high threshold π^R of the average saturation of label probabilities in the R-step should be used, where saturation was defined as a single label's probability approaching 1 while the rest are close to 0. This condition will occur when the compatibility functions r (used in Eqn 5.15) for the label return very low values and as $\beta_z \to \infty$ in Eqn 5.14. The square of the label probabilities penalised large deviations

from 1 harshly to ensure that almost all of the sub-voxel labels were highly saturated when convergence was reached.

Definition 5.3.2 (Labelling saturation) The labelling of a volume is considered **saturated** at sub-iteration z of the R-step if for each sub-voxel $i \in N$ the square of the neighbourhood label context prior (Eqn 5.14) summed over all labels $m \in M$ matches or exceeds a threshold $0 \ll \pi^R < 1$.

$$\frac{\sum_{i=1}^{N} \sum_{m \in M} p_z(m|\phi, p(f_i^{z-1}))^2}{N} \geq \pi^R \tag{5.23}$$

Convergence of the main EM iterations can be checked by investigating the absolute difference between iterations $(t-1)$ and t in the labelling F after the E-step (Dfn 5.3.3); the labelling should differ by only a very small amount in the final iteration. $0 \leq \pi^E \ll 1$ defines the average labelling change threshold which must be surpassed for processing to continue.

Definition 5.3.3 (Expectation-maximisation convergence) The expectation maximisation algorithm is considered to have reached convergence when the mean absolute difference of the labelling F over all classes C between iterations $(t-1)$ and t falls below a threshold $0 \leq \pi^E \ll 1$. This threshold is effectively doubled since the maximum possible difference in the labelling of a voxel indexed by $i \in N$ is 2.

$$\frac{\sum_{i=1}^{N} \sum_{k \in C} \left| F_{ik}^{t} - F_{ik}^{t-1} \right|}{2N} \leq \pi^E \tag{5.24}$$

5.3.7 Processing steps and algorithm summaries

Procedure 5.3.1 (Pre-processing) The following pre-processing steps must take place before the initialisation (Proc 5.3.2) and main segmentation (Proc 4.2.2) stages can be performed:

1. The T_2- and PD-weighted volumes must first undergo skull stripping to produce a parenchymal mask, as outlined in Section 2.2;
2. The thin structure map B (Dfn 3.4.5), the shaped and oriented anisotropic filter set and the filter indices ϕ for sub-voxels flagged in B are computed, as described in Chapter 3;
3. The volumes' resolution is enhanced by ratio L, as detailed in Chapter 4.

Procedure 5.3.2 (Initialisation) After pre-processing (Proc 5.3.1), the following initialisation steps must take place before the main segmentation stage (Proc 5.3.3) can be performed:

1. The label set M allowed at the sub-voxel level is generated using the fractions specified in M^{frac} for a constrained mixel dimension of 2, with respect to any disallowed class mixtures or mixing fractions;
2. The initial class parameters $\bar{\theta}^0$ are derived from samples of BG intensities selected from the corners of the volume as well as from likely pure samples of the other classes taken from thresholded memberships of the other classes generated by FCM clustering, as described in Section 5.3.2;
3. Constant priors for thin structure localisation (Dfn 5.3.1) and enforcing mixel dimension constraints (Eqn 5.19) are then computed using M, B and L.

Procedure 5.3.3 (Main segmentation algorithm) Once pre-processing and initialisation (Procs 5.3.1 and 5.3.2) have taken place, the main segmentation stage can be performed. Each iteration consists of the following steps:

1. The conditional probabilities of the sub-voxel intensities (Eqn 5.12) are calculated using the current estimate of the pure tissue class parameters $\overline{\theta}$;
2. Sub-iterations of the R-step (described in Section 5.3.4) are performed until the label probabilities reach saturation (Eqn 5.3.2). After each iteration indexed by z, the value of the ambiguity reduction schedule β_z (used in Eqn 5.14) is reduced;
3. The convergence of the EM algorithm (Dfn 5.3.3) is checked after completion of the E- and R-steps, so that the algorithm can terminate;
4. The sub-voxel labels and intensities are downsampled to the original resolution (using Eqn 5.5), and returned if convergence has been reached;
5. Only spatially homogeneous labels are selected as samples for the re-estimation of pure tissue class parameters $\overline{\theta}$ in the M-step, as detailed in Section 5.3.5.

5.4 Demonstrations

The initial investigations in this section attempt to quantify the performance gains of the methods described in Section 5.3 in a progressive fashion by sequentially incorporating the methods associated with producing resolution-enhanced data (in Section 5.4.1), relaxation labelling (Section 5.4.2), and thin structure priors as well as the shaped and oriented neighbourhood weighting changes (both in Section 5.4.4) into basic EM-based soft segmentation. The M-step (Section 5.3.5) and the conditional probability of the intensities (Eqn 5.12) were applied in their entirety, with the greedy ICM [Besag, 1986] algorithm optimising the MRF in place of the R-step (Section 5.3.4) where applicable. In these initial sections the effects of changing the sub-voxel level label set M have also been examined (Section 5.4.3) in order to design an appropriate set for further testing of the entire system. ROI 5 has again been selected for the

investigation of methodological and parameter changes due to its inclusion of several different anatomical structure types including blood vessels, cortical sulci, convoluted cortex, deep GM nuclei, the skull, the edge of the lateral ventricle and thin regions of WM. The absolute error, κ statistic scores and volume errors calculated on volume-wide and per-class bases for each of the following four subsections are found in Table 5.3.

Later testing phases focus on quantification of the complete system's accuracy over ROIs taken from the LBC1936 and BrainWeb datasets where the ground truth has been made available. Comparisons with the competing segmentation method FAST [Zhang et al., 2001] were performed as detailed in the initial investigations of its accuracy in Section 1.3.5. Other parameter values for generation of the thin structure maps and the shaped and oriented filters (listed in Section 3.5) and resolution enhancement of the volumes (Section 4.3) remained the same as those employed in other chapters. The remaining parameters for the proposed segmentation process were set as follows, unless otherwise stated in each section:

- The saturation threshold $\pi^R = 0.975$ for relaxation labelling (used in Dfn 5.3.2) was chosen to be close to unity, to ensure almost all sub-voxels had reached a highly saturated labelling state by the end of the R-step;
- $\beta_z = \frac{\beta_{z-1}}{2}$ with initial condition $\beta_0 = 0.3$ was selected to allow some ambiguity in the labelling at the start of the R-step but also so that there were not too many opportunities for label interactions that could lead to an over-smoothed result;
- The tissue prior penalty $pen = 0.5$ was set high enough to ensure the labelling could benefit from the high PPV offered by the binary thin structure maps B (Dfn 3.4.5) to help disambiguate mixture components, but would still not be substantially affected by any errors in B;
- The constrained mixel dimension penalty $cmdpen = 1.5$ was set so the constrained mixel dimension was relevant in early (higher labelling ambiguity) R-step sub-iterations where the establishment of the correct constituent classes in each voxel was most important;
- The basic set of fractional levels allowed for each tissue class were $M^{frac} = \{0, 0.5, 1\}$, but BG/WM and BG/GM mixtures were banned. Furthermore, WM/GM mixtures could also take the 0.35/0.65 level. The motivation for these restrictions is explained in Section 5.4.3;

- Initialisation was provided by FCM[11] with $\alpha = 0.7$ and $p = 2$. The initial centroid values $V_{T2}^0 = [0.17, 0.25, 0.85]$ and $V_{PD}^0 = \{0.35, 0.45, 0.75\}$ for the WM, GM and CSF classes were chosen by manual sampling of pure tissue regions in frontal WM, pre-frontal cortical gyri and the ventricles respectively from normalised volumes in the LBC1936 dataset;

- Both the isotropic and anisotropic (shaped and oriented) neighbourhood weighting filters were Gaussian with an applicability range of $5 \times 5 \times 5$ sub-voxels. The isotropic filter had a spread $\sigma = L$, and like Section 3.5 the shaped and oriented filters possessed a *large* : *small* ratio of 2:1 with Σ set to preserve the volume of the spheroid isotropic filter in the ellipsoid anisotropic ones;

- The threshold $\pi^E = 0.002$ for labelling change between E-step iterations (Dfn 5.3.3) was chosen so that the EM algorithm could terminate within a few iterations.

5.4.1 Effects of using resolution enhanced data

Even when processed without using thin-structure-specific processing such as changes to the local levels, resolution-enhanced data still offers two advantages over the original data: low magnitude noise is smoothed and partial volume voxels at region boundaries are decomposed into sub-voxels for which a majority will possess intensities similar to each region. Some differences in labelling resolution-enhanced data using FCM clustering have already been highlighted (Fig 4.15) but in this section the EM-based algorithm proposed in Section 5.3 will be applied. However, the algorithm was stripped down to make it more similar to most typical segmentation algorithms; the greedy ICM [Besag, 1986] algorithm was used in place of relaxation labelling to optimise the labelling in the context of neighbours, and no shaped and oriented neighbourhood weighting nor thin structure localisation priors were used.

The most striking change when performing segmentation on resolution enhanced data appeared to be the increased coverage of cortical sulci, evident as reduced GM content and similarly increased CSF in Fig 5.4E. At the sub-voxel scale sulci are effectively up to double the thickness (since $L = [2, 2, 2]$) in terms of elements spanned, and so more neighbours within the sulci can occupy the greatest neighbourhood filter weights. However, pure CSF labelling is still not as prevalent as in the ground truth

[11] These values were recommended in [Ahmed et al., 2002] for MRI volumes with moderate amounts of noise.

Basic parameter set/methodology														
	Absolute error					κ-*statistic score*					*Volume error*			
Section	**BG**	**WM**	**GM**	**CSF**	**All**	**BG**	**WM**	**GM**	**CSF**	**All**	**BG**	**WM**	**GM**	**CSF**
5.4.1	15.5	399	505.5	130	1050	0.267	0.515	0.460	0.705	0.502	-15.5	258	-134.5	-108
5.4.2	18	386.5	437	111.5	953	0.328	0.510	0.489	0.762	0.529	-10	267.5	-216	-41.5
5.4.3	17.5	380	459.5	122	979	0.213	0.563	0.535	0.730	0.564	-15.5	245	-171.5	-58
5.4.4	17.5	324.5	389	107	838	0.253	0.570	0.548	0.774	0.584	-14.5	180.5	-134	-32
Modified parameter set/methodology														
5.4.1	18	386.5	437	111.5	953	0.328	0.510	0.489	0.762	0.529	-10	267.5	-216	-41.5
5.4.2	17.5	324.5	389	107	838	0.253	0.570	0.548	0.774	0.584	-14.5	180.5	-134	-32
5.4.3	18.5	406	497	122.5	1044	0.190	0.396	0.381	0.731	0.440	-13.5	-102	177	-61.5
5.4.4	15	319	381.5	101.5	817	0.370	0.554	0.542	0.785	0.579	-10	117	-74.5	-32.5

Table 5.3: **Error statistics for parameter and methodology changes over ROI 5 examined in Sections 5.4.1-5.4.4**. Errors from the expert labelling are listed to 3 decimal places. The best error statistics for each test between the different parameter sets or methodologies is hightlighted in red.

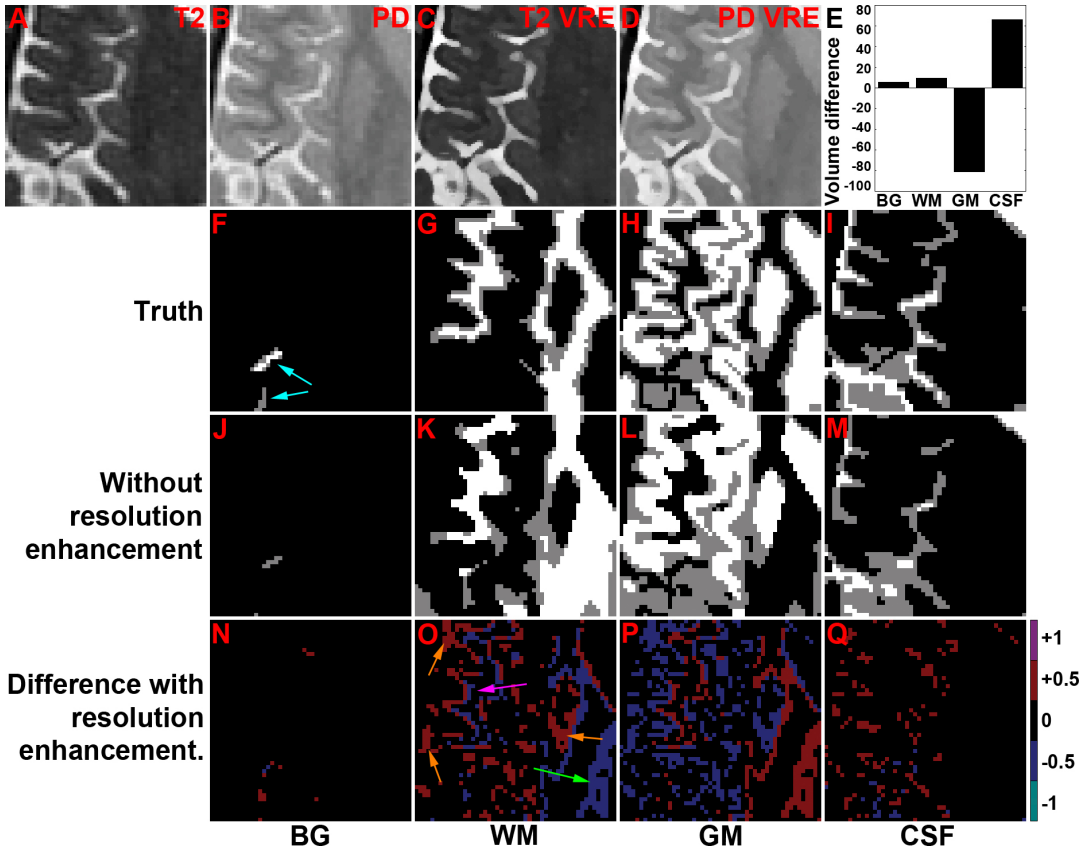

Figure 5.4: **Improvements to segmentation quality by resolution enhancement of the data**. Processing methods detailed in Section 5.4.1. **A** & **B**) T_2- and PD-weighted data for ROI 5. **C** & **D**) Resolution enhancement of **A** and **B**, without shaped and oriented filtering nor thin structure restoration. **res. enh.**: resolution enhanced. **E**) Graph of class volume differences for the processing using the resolution enhanced data, compared to using only the original data. **F** – **I**) Expert labelling of each tissue class; white is pure tissue, gray is a partial volume, and black denotes no tissue content. **J** – **M**) Segmentation result using the original data only; white denotes pure tissue content, black denotes no tissue content, and shades of gray denote fractional volume content. **N** - **Q**) Labelling differences from **J** – **M** using the resolution enhanced data. Coloured arrows are referenced in Section 5.4.1.

as no thin-structure-specific processing has been performed to restore pure intensities. GM content also changes at WM/GM borders due to the increased sensitivity to thin regions of WM between cortical gyri (marked with a pink arrow) as well as improvements in the congruence of the borders of poorly-contrasted deep GM nuclei (green arrows) between the different image weighting factors. However, WM/GM partial volume is erroneously labelled in some thin cortical regions and between deep GM nuclei (orange arrows) due to smoothing around poorly-contrasted boundaries.

BG experiences a little decrease in accuracy overall, mainly due to erroneous labelling of strong dark noise that was not smoothed by resolution enhancement. Since label optimisation was performed using ICM [Besag, 1986], then the dark intensities of both blood vessels (aqua arrows) were initially assigned WM- or GM-based labels from the results of the conditional probabilities of the intensities (Eqn 5.12) and a BG-based label did not enter consideration. Nonetheless, improvements in both error metrics were observed for all the other classes, leading to an overall 10% decrease in absolute error and 0.027 increase in the κ-statistic score.

5.4.2 Impact of performing neighbourhood optimisation using relaxation labelling

The testing in the previous section employed a simplified EM-based segmentation method aiming to replicate some of the properties of commonly-used segmentation algorithms. This strategy is repeated – including the use of resolution-enhanced data – to investigate the effects of relaxation labelling (described in Section 5.3.4). This label optimisation method replaced ICM [Besag, 1986] in the modified methods, and it is expected that benefits in mixture disambiguation and poorly-contrasted class discrimination should be made apparent due to its consideration of all labels, rather than just the most likely one in each iteration of label optimisation. Indeed, this was the case: overall there was a 12% decrease in absolute error and 0.055 improvement in the κ-statistic score, with substantial improvements in both metrics for WM and GM.

The means of WM and GM are not widely separated in T_2-PD volumes (as illustrated in Fig 1.11) and so the relaxation process helped to correct cortical GM identified as at least partial WM (marked with aqua arrows in Fig 5.5) and classified more pure GM at the core of the putamen (pink arrow). A net loss in WM was matched with a very similar gain in GM (Fig 5.5E) but there were few substantial changes to CSF, which has a bright intensity and hence its mixtures are more easily disambiguated.

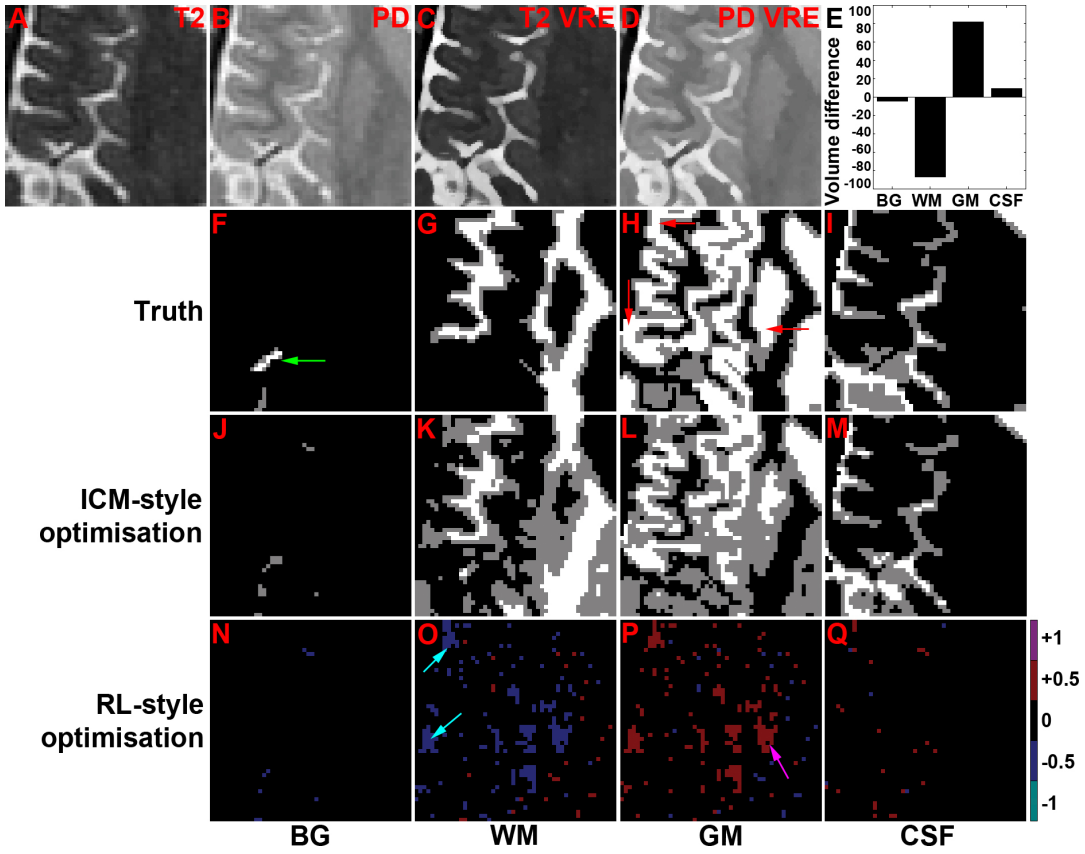

Figure 5.5: **Effect of changing the neighbourhood optimisation method**. Processing methods detailed in Section 5.4.2. **A** & **B**) T_2- and PD-weighted data for ROI 5. **C** & **D**) Resolution enhancement of **A** and **B**, without shaped and oriented filtering nor thin structure restoration. **res. enh.**: resolution enhanced. **E**) Graph of class volume differences for the processing using relaxation labelling, compared to using only ICM. **F** – **I**) Expert labelling of each tissue class; white is pure tissue, gray is a partial volume, and black denotes no tissue content. **J** – **M**) Segmentation result using the original data only; white denotes pure tissue content, black denotes no tissue content, and shades of gray denote fractional volume content. **N** - **Q**) Labelling differences from **J** – **M** using relaxation labelling. Coloured arrows are referenced in Section 5.4.2.

Some noisy voxels were removed from the BG labelling but there was no significant increase in accuracy; the topmost vessel (green arrow) was less fully labelled due to the poor representation of pure BG intensities in the neighbourhood, to which the clique potential scoring function ζ (Eqn 5.18) of relaxation labelling appears to be sensitive.

5.4.3 Differences in performance when changing the sub-voxel label set

In this section the putative benefits of improved sensitivity associated with segmentation performed using a denser label set ($M^{frac} = \{0, 0.35, 0.5, 0.65, 1\}$) at the upscaled level will be compared against a minimal, and likely more robust set ($M^{frac} = \{0, 0.5, 1\}$). Several class mixtures have been identified in Section 5.4 as unlikely to occur within volumes acquired at typical clinical resolutions, notably BG/WM and BG/GM, and so these were excluded from the latter set. Like the previous sections the testing will be performed using a stripped-down EM-based segmentation algorithm, although this time resolution-enhanced data and relaxation labelling will be included in processing as some benefits have been established. The effects of implementing thin-structure-specific processing will be assessed in the following section once an appropriate label set has been designed.

The absolute errors and κ-statistic scores for the minimal label set displayed either the same values or improvements over those for the larger set. WM labelling was particularly poor when extra fractional levels were available, causing all the pure WM regions to be given a 0.35/0.65 WM/GM label (marked with an aqua arrow in Fig 5.6). However, the smaller label set does lack expressiveness for some true WM/GM mixing (pink arrow) and the extent of some sulci is a little fuller using the larger set (green arrow). The improvements to labelling sulci may not be apparent, however, when pure CSF intensity is restored in the next section. These findings prompt the design of a compromise label set, with at its heart the minimal and constrained approach. An additional high WM/low GM (0.65/0.35) label, that should not cause confusion in pure WM regions, will allow better labelling of some WM/GM boundaries and the deep GM nuclei.

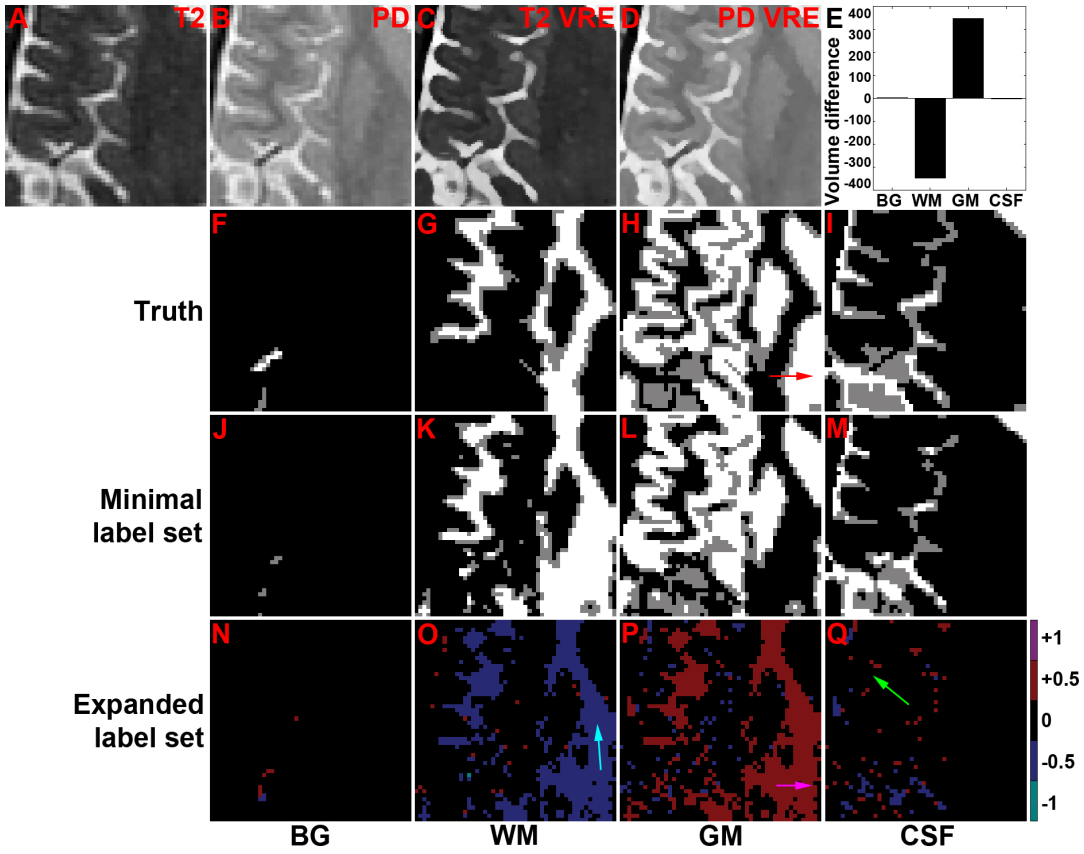

Figure 5.6: **Effect of varying the range of fractional tissue levels allowed on segmentation quality**. Processing methods detailed in Section 5.4.3. **A** & **B**) T_2- and PD-weighted data for ROI 5. **C** & **D**) Resolution enhancement of **A** and **B**, without shaped and oriented filtering or thin structure restoration. **res. enh.**: resolution enhanced. **E**) Graph of class volume differences for the processing using the expanded labels set, compared to using only the minimal label set. **F** – **I**) Expert labelling of each tissue class; white is pure tissue, gray is a partial volume, and black denotes no tissue content. **J** – **M**) Segmentation result using the original data only; white denotes pure tissue content, black denotes no tissue content, and shades of gray denote fractional volume content. **N** - **Q**) Labelling differences from **J** – **M** using the expanded label set. Coloured arrows are referenced in Section 5.4.3.

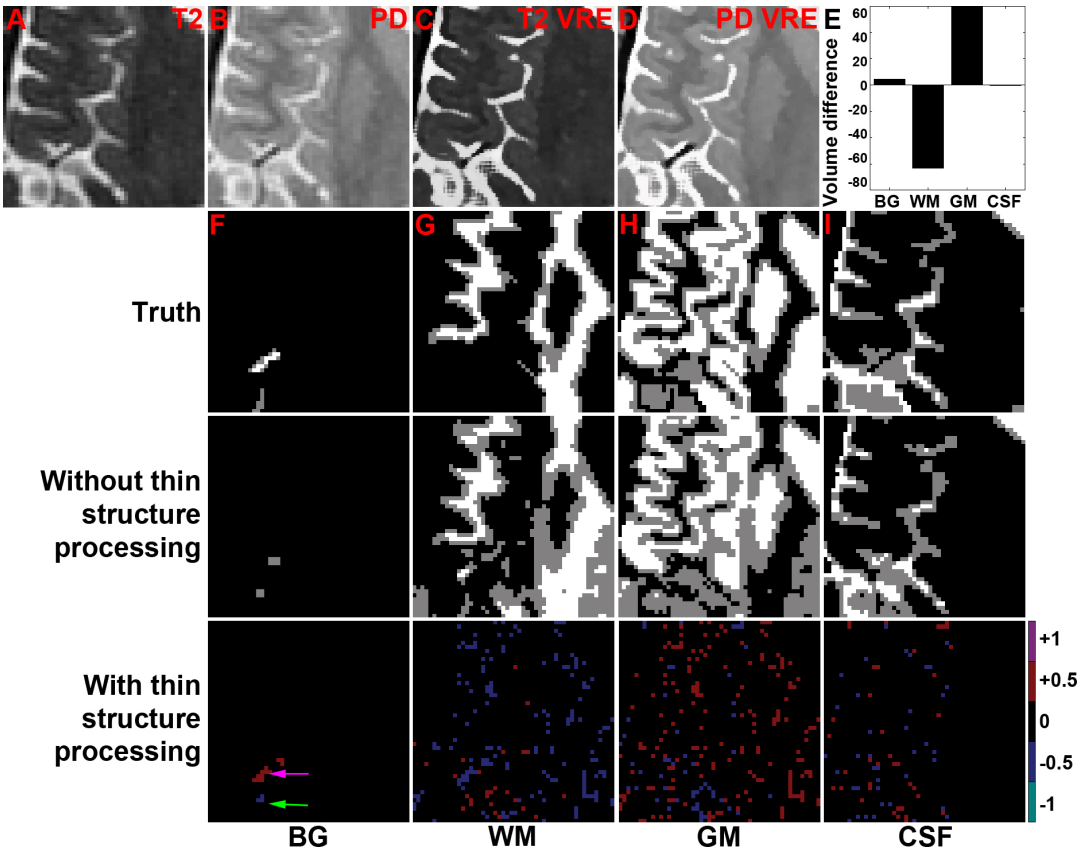

Figure 5.7: **Improvements to segmentation quality by applying thin structure priors and shaped and oriented filtering**. Processing methods detailed in Section 5.4.4. **A** & **B**) T_2- and PD-weighted data for ROI 5. **C** & **D**) Resolution enhancement of **A** and **B**, with shaped and oriented filtering and thin structure restoration. **res. enh.**: resolution enhanced. **E**) Graph of class volume differences for the processing using the thin structure priors and shaped and oriented filtering, compared to not using them. **F** – **I**) Expert labelling of each tissue class; white is pure tissue, gray is a partial volume, and black denotes no tissue content. **J** – **M**) Segmentation result using the original data only; white denotes pure tissue content, black denotes no tissue content, and shades of gray denote fractional volume content. **N** - **Q**) Labelling differences from **J** – **M** using thin structure priors and shaped and oriented filtering. Coloured arrows are referenced in Section 5.4.4.

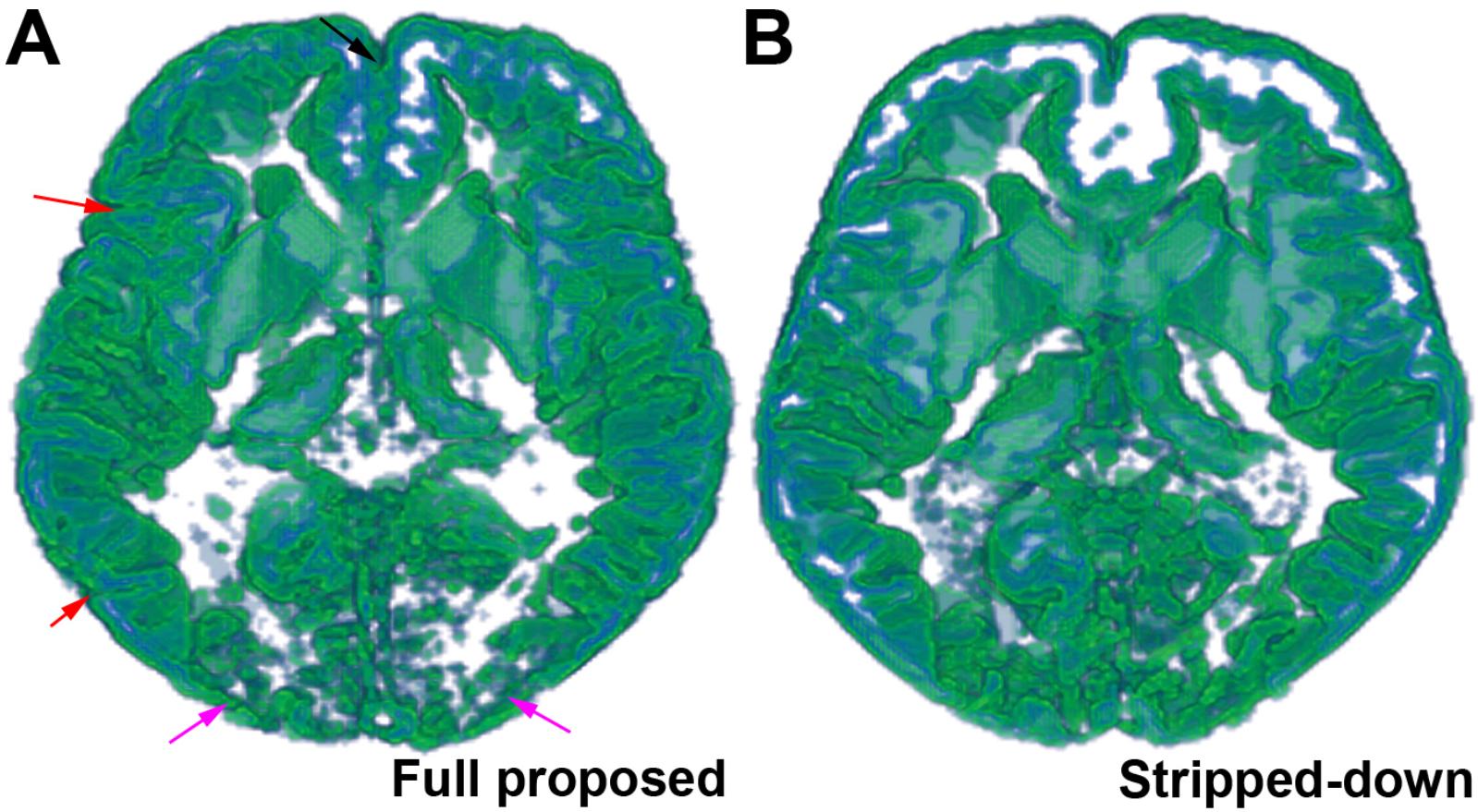

Figure 5.8: **Visualisation of differences in estimated GM thickness by applying the proposed methods**. Simian rendering of the GM class with thin structure priors, shaped and oriented filtering and resolution enhancement of the volume data (**A**) and without (**B**) shown. Source data was a sub-volume of 15 slices surrounding ROI 5.

5.4.4 Incorporation of thin structure priors and shaped and oriented filtering

Previous testing in this chapter has sequentially presented the benefits of using resolution-enhanced data (Section 5.4.1), relaxation labelling (Section 5.4.2) and a fairly minimally-sized label set (Section 5.4.3). While the former two modifications have been used previously in the segmentation of brain MRI volumes, thin-structure-specific processing for both the segmentation and resolution enhancement stages has not been considered in the literature. The effects of using shaped and oriented filtering, adaptive neighbourhood weighting as well as thin structure localisation priors have been assessed last, so that their benefits can be highlighted in the context of the established previous changes.

Since the means of WM and GM tend to be poorly separated in T_2-PD volumes, even slight changes to the labelled thickness of the cortex – coming about through thin-structure-specific processing of sulci and vessels – can manifest in significant changes to the updated class parameters, and in turn the localisation of WM/GM boundaries (Fig 5.7). The extreme contrasts of BG and CSF ensure that there are not quite as drastic effects for these classes due to this factor, which instead enjoy improvements in their κ-statistic scores. Overall there is a 3% improvement in absolute error with reductions in error for all classes.

Closer examination of the thin structures revealed the extent of the topmost vessel (pink arrow) is much fuller, although pure content at its core is still not labelled. The pure BG intensities are visible in the resolution enhanced data but the high slice thickness causes this vessel to remain poorly-resolved at the sub-voxel level, leading to only partial volume labels being formed when downsampling. There is no substantial change for the vessel below it (green arrow) which is not highlighted in B^{BG} and thus remains blurred in the resolution-enhanced data. The sulci in this ROI are fairly well resolved and so differences only manifest as small changes at their borders with GM.

A rendering of GM content over 15 slices surrounding ROI 5 (Fig 5.8), comparing the stripped-down segmentation method first investigated in Section 5.4.1 with the full system, showed a much greater sensitivity to small detail. There were visible cortical invaginations caused by sulci (marked with red arrows), thin regions of WM and the falx cerebri (black arrow), and the cortical thickness remained similar in most regions. However, in occipital regions (pink arrows) weaknesses in enhancing WM/GM boundaries caused a reduction in the observed thickness.

5.4.5 Comparative summary of performance gains by incorporating the methods proposed in this thesis

A round-up of some of the effects on segmentation quality of incorporating subsets of the methods described in this thesis is presented in Fig 5.9, building up from an algorithm using no anatomical structure priors nor adaptive filters, as well as no resolution enhancement, and ICM-based label optimisation in place of relaxation labelling. In this first case (panels E-H) the prevalence of pure tissue labelling is high due to the design of the clique potential function (Eqn 5.18), which expects mostly pure tissue to exist at the sub-voxel level used in the full proposed methods. Consequently some soft WM/GM boundaries are incorrectly labelled (marked with a blue arrow), and both BG and CSF content is poorly estimated due to it mainly existing as partial volumes. Adding volume resolution enhancement (panels I-L) and processing at the sub-voxel level remedies these issues somewhat but there is some underestimation of BG (purple arrow) and overestimation of CSF (orange arrow). In turn, the correction of these mistakes is tackled by adding thin structure detection into the processing pipeline (panels M-P and Q-T).

Panels Q-T and U-X replaced the ICM algorithm, used to optimise the labelling in the context of neighbours, with relaxation labelling. The main differences shown are the greater extent of labelling of the thalamus (green arrow), improvements in the uniformity of pure GM labelling in the cortex (cyan arrow), and enhanced sensitivity to structure in regions containing CSF (yellow arrow).

5.4.6 LBC1936 dataset

The performance of FAST [Zhang et al., 2001] in labelling real brain ROIs of the LBC1936 dataset has already been investigated in Section 1.3.5 as motivation for work in this thesis, but this section undertakes a comparative evaluation of the effectiveness of the improvements. Quantitative absolute error and κ-statistic score metrics (presented in Table 5.4 in all-class and per-class formats) have been analysed, and qualitative assessment of the class content maps (pictured in Figs 5.10-5.15) has also been performed. The latter analyses have only been performed for ROIs 1-6 since ROIs 7-10 present similar anatomical regions.

GM content was always better estimated using the proposed method, with on average a 27% reduction in absolute error. The most striking difference occurs in ROI 4 (Fig 5.13) where erroneous labelling of pure GM in the cerebellum by FAST (marked

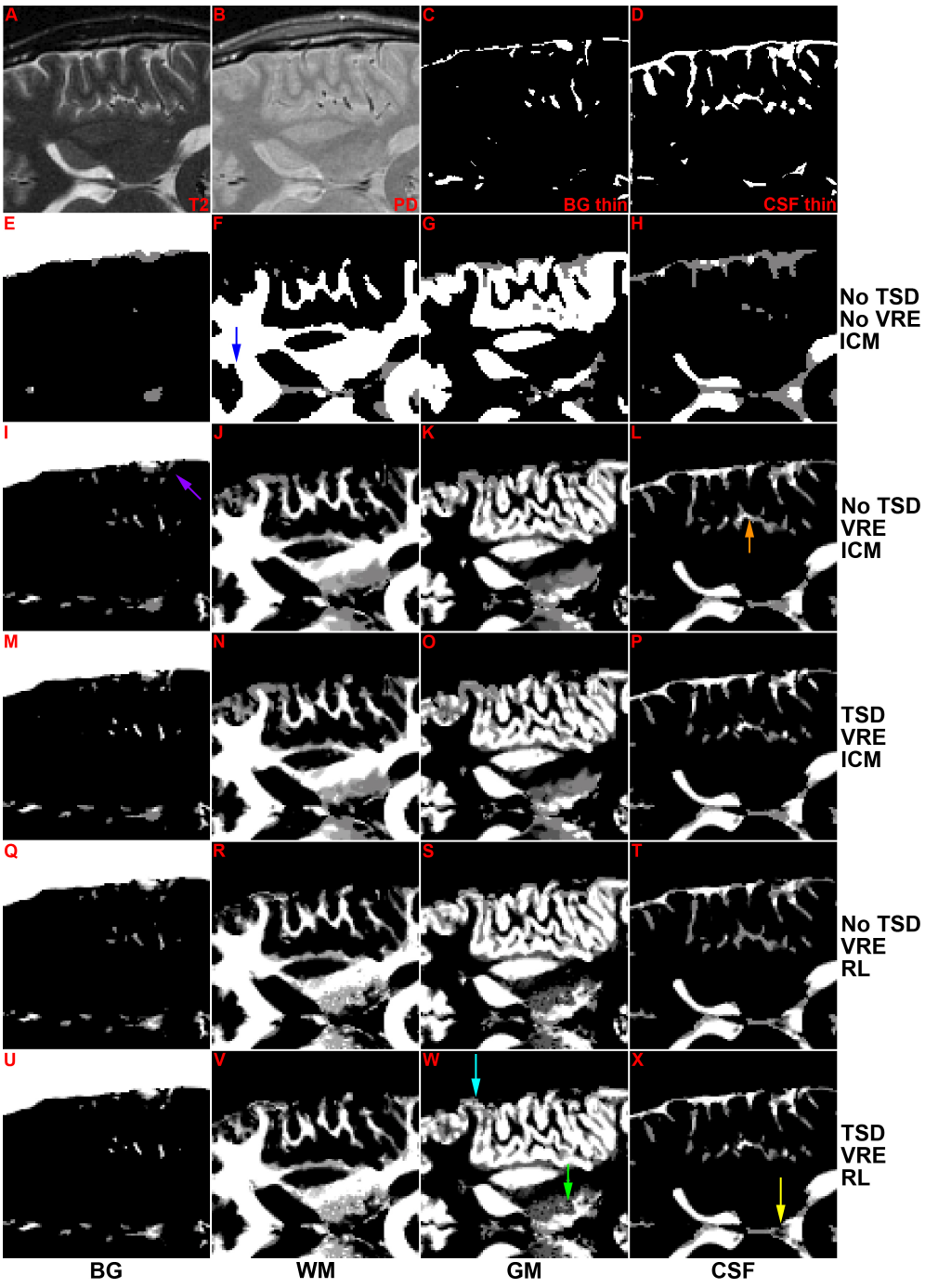

Figure 5.9: **Comparison of segmentation results over a region from the young brain dataset, with subsets of the methods described in this thesis disabled. A** & **B**) T_2- and PD-weighted data. **C** & **D**) Thin anatomical structure prior maps for BG and CSF. **E**-**X**) Segmentation results when using subsets of the methods (shown in the final row). **VRE**: volume resolution enhancement used. **ICM**: iterated conditional modes optimisation used. **RL**: relaxation labelling optimisation used. **TSD**: thin structure detection used. Coloured arrows are referenced in Section 5.4.5.

FAST														
	Absolute error					*κ-statistic score*					*Volume error*			
ROI	**BG**	**WM**	**GM**	**CSF**	**All**	**BG**	**WM**	**GM**	**CSF**	**All**	**BG**	**WM**	**GM**	**CSF**
1	17	254.5	336	148.5	756	0.257	0.668	0.540	0.745	0.617	15	-102.5	6	81.5
2	31	401.5	545.5	265	1243	0.000	0.463	0.309	0.492	0.360	31	206.5	-223.5	-14
3	41.5	432.5	628	328	1430	0.189	0.297	0.249	0.469	0.279	-9.5	-44.5	-181	235
4	0	544.5	837	461.5	1843	0.000	0.165	0.064	0.403	0.164	0	-395.5	401	-5.5
5	16	332	466	156	970	0.383	0.577	0.500	0.702	0.547	-6	94	-193	105
6	7	639	708	110	1464	0.000	0.291	0.314	0.820	0.420	7	464	-520	49
7	52.5	527.5	513.5	96.5	1190	0.116	0.438	0.380	0.801	0.483	50.5	-476.5	425.5	0.5
8	4.5	461.5	499.5	49.5	1015	0.248	0.282	0.348	0.767	0.370	2.5	55.5	-76.5	18.5
9	14	152	285	169	620	0.400	0.774	0.594	0.649	0.639	6	-29	-128	151
10	14	361.5	492	227.5	1095	0.778	0.548	0.344	0.474	0.404	11	304.5	-257	-58.5
Proposed method														
1	6.5	225.5	295	114	641	0.433	0.716	0.607	0.773	0.671	0.5	-17.5	85	-68
2	47	566.5	467.5	201	1282	0.000	0.313	0.359	0.574	0.325	47	494.5	-369.5	-172
3	32	430	540.5	195.5	1198	0.506	0.234	0.284	0.642	0.351	3	-116	133.5	-20.5
4	20.5	372.5	250	286	929	0.000	0.472	0.224	0.577	0.431	20.5	296.5	-113	-204
5	15	319	381.5	101.5	817	0.370	0.554	0.542	0.785	0.579	-10	117	-74.5	-32.5
6	0.5	453.5	488	132	1074	0.000	0.387	0.492	0.774	0.515	0.5	257.5	-135	-123
7	7	254.5	287	84.5	633	0.515	0.662	0.531	0.819	0.645	7	-132.5	200	-74.5
8	1.5	463	476	44.5	985	0.666	0.326	0.410	0.767	0.413	1.5	88	-52	-37.5
9	14.5	176	241	88.5	520	0.329	0.759	0.681	0.780	0.704	5.5	0	64	-69.5
10	69.5	479	445.5	134	1128	0.373	0.394	0.334	0.679	0.371	69.5	387	-339.5	-117

Table 5.4: **Error statistics for the proposed segmentation method and FAST over expert-labelled ROIs from the LBC1936 dataset**. Errors are listed to 3 decimal places. The smallest absolute error and largest κ statistics between the proposed method and FAST are highlighted in red; equal values are highlighted in blue. 2500 voxels were present in each ROI.

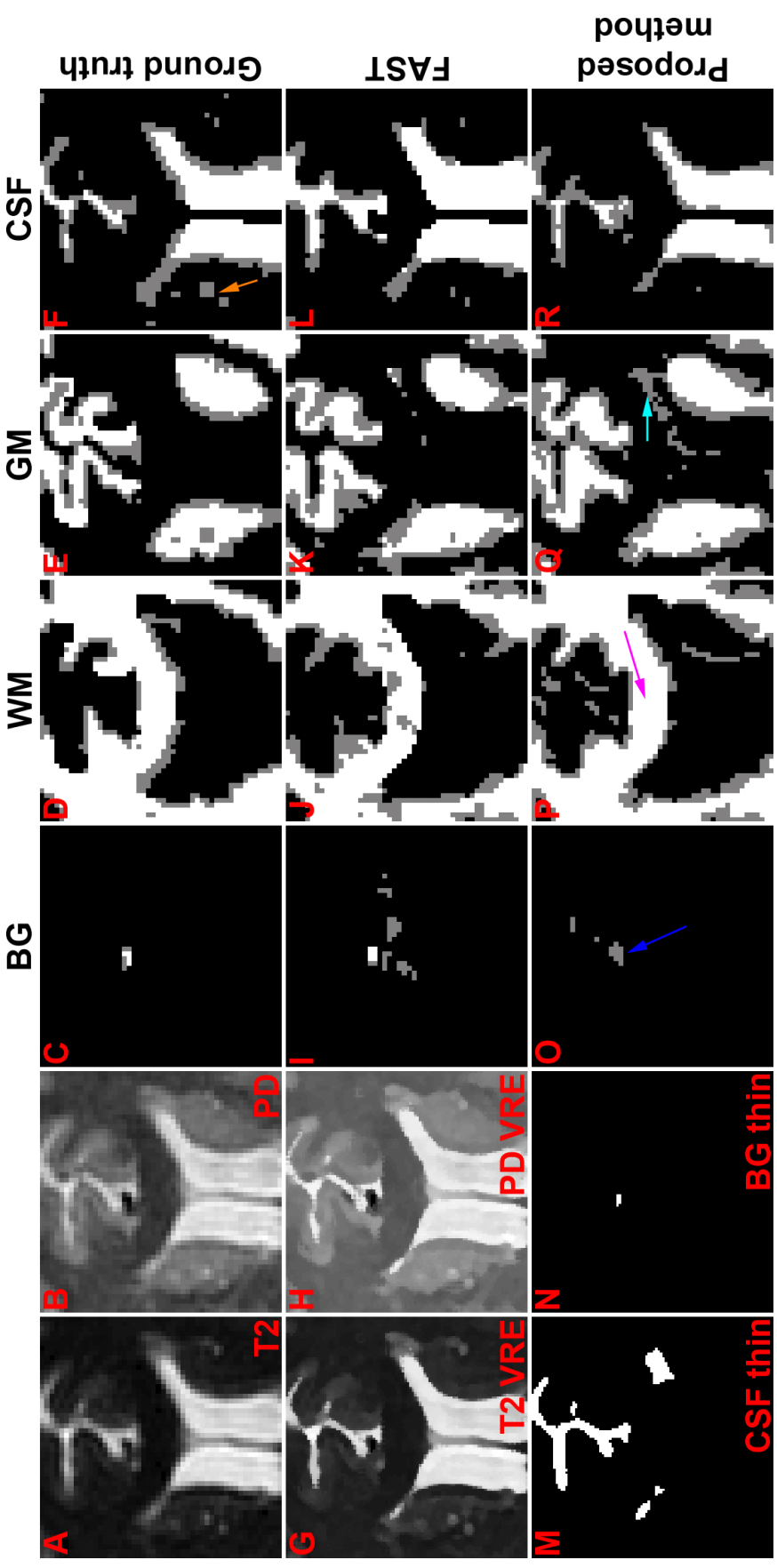

Figure 5.10: **Segmentation of ROI 1**. **A** & **B**) T_2- and PD-weighted data. **C** - **F**) Expert labelling of each tissue class. **G** & **H**) Resolution-enhanced T_2- and PD-weighted data. **I** - **L**) Labelling of each tissue class produced by FAST. **M** & **N**) Binary thin CSF and BG structure maps. **O** - **R**) Labelling of each tissue class produced by the proposed methods. Coloured arrows are referenced in Section 5.4.6.

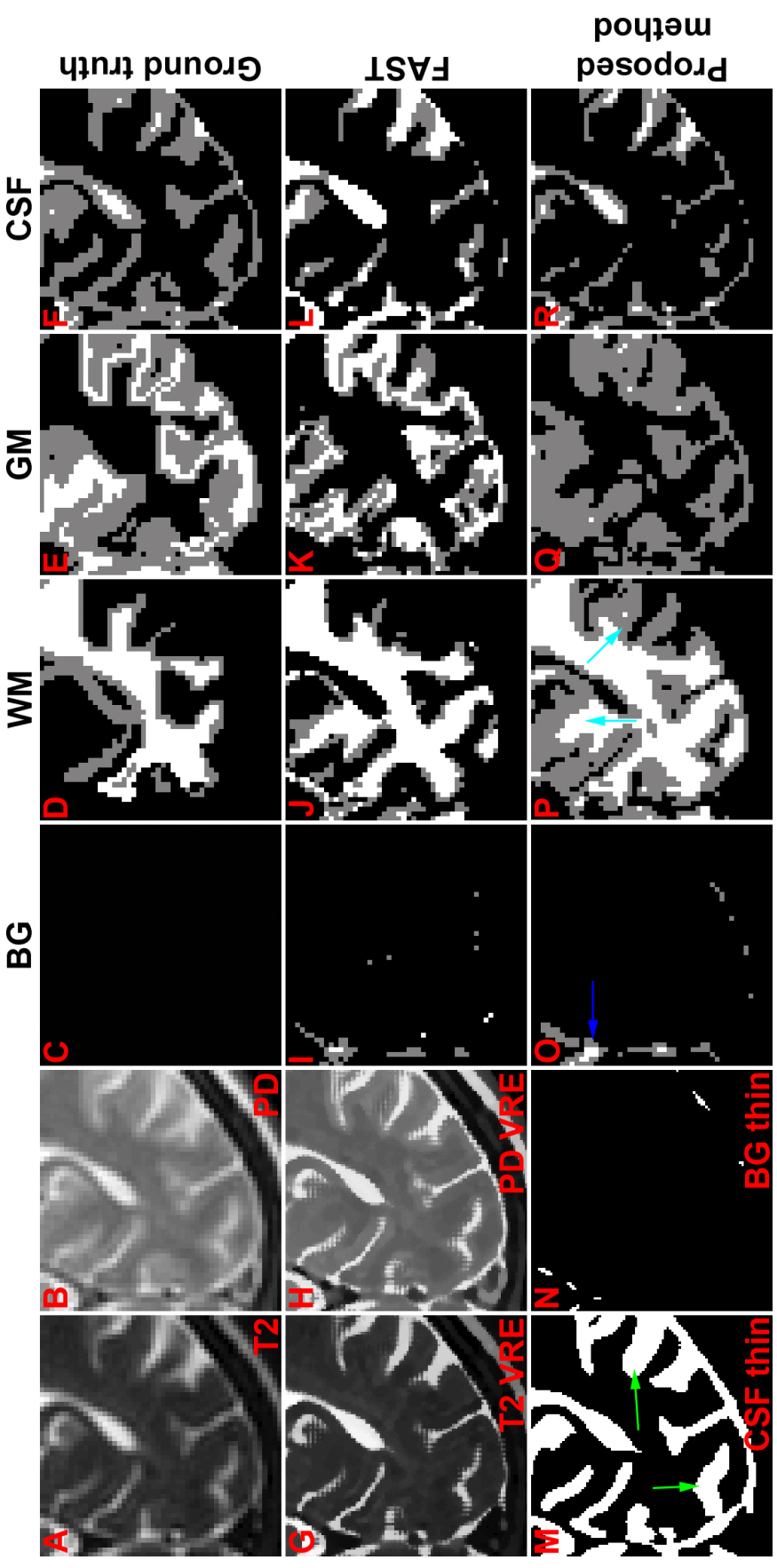

Figure 5.11: **Segmentation of ROI 2**. **A** & **B**) T_2- and PD-weighted data. **C** - **F**) Expert labelling of each tissue class. **G** & **H**) Resolution-enhanced T_2- and PD-weighted data. **I** - **L**) Labelling of each tissue class produced by FAST. **M** & **N**) Binary thin CSF and BG structure maps. **O** - **R**) Labelling of each tissue class produced by the proposed methods. Coloured arrows are referenced in Section 5.4.6.

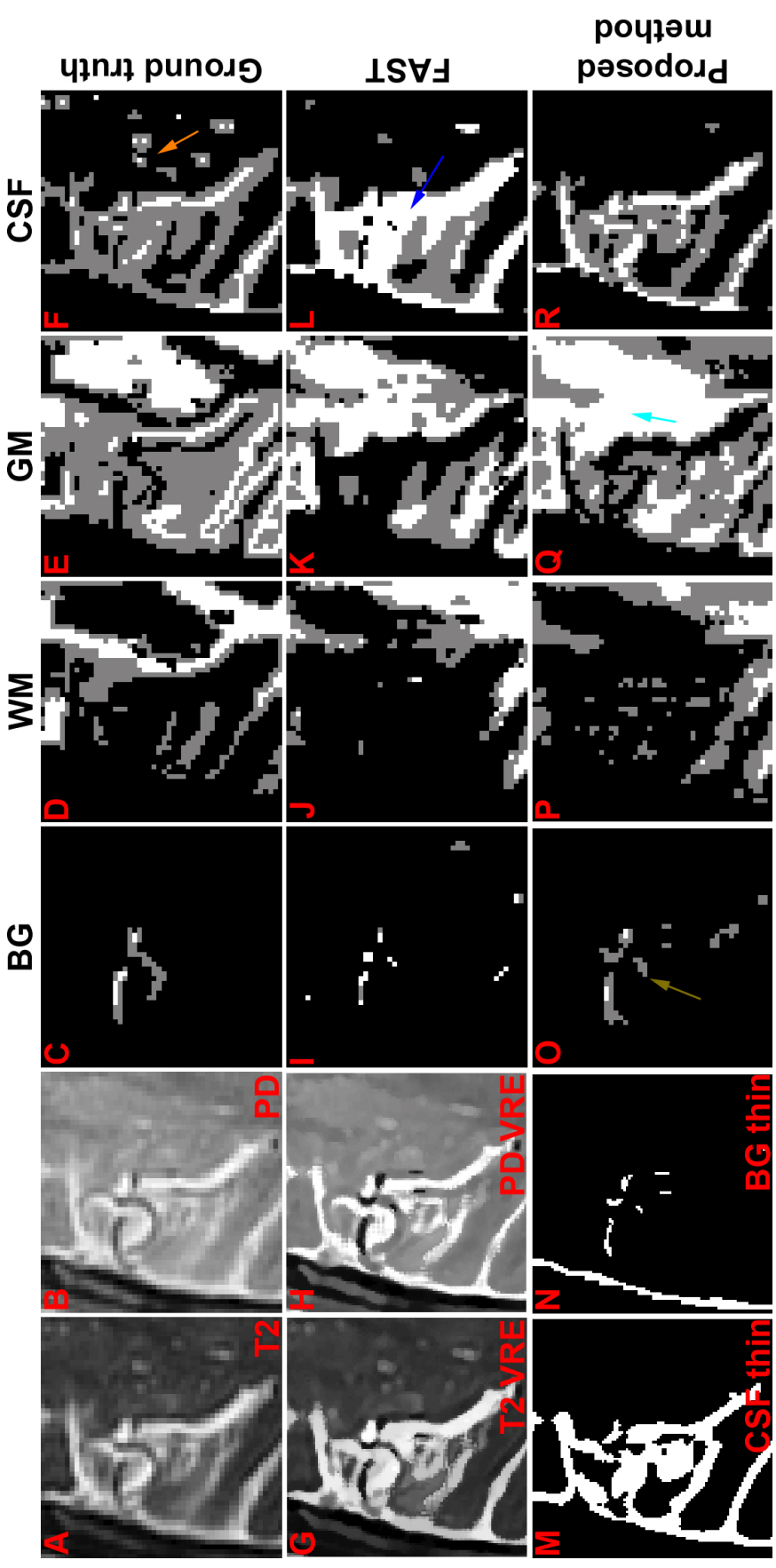

Figure 5.12: **Segmentation of ROI 3**. **A** & **B**) T_2- and PD-weighted data. **C** - **F**) Expert labelling of each tissue class. **G** & **H**) Resolution-enhanced T_2- and PD-weighted data. **I** - **L**) Labelling of each tissue class produced by FAST. **M** & **N**) Binary thin CSF and BG structure maps. **O** - **R**) Labelling of each tissue class produced by the proposed methods. Coloured arrows are referenced in Section 5.4.6.

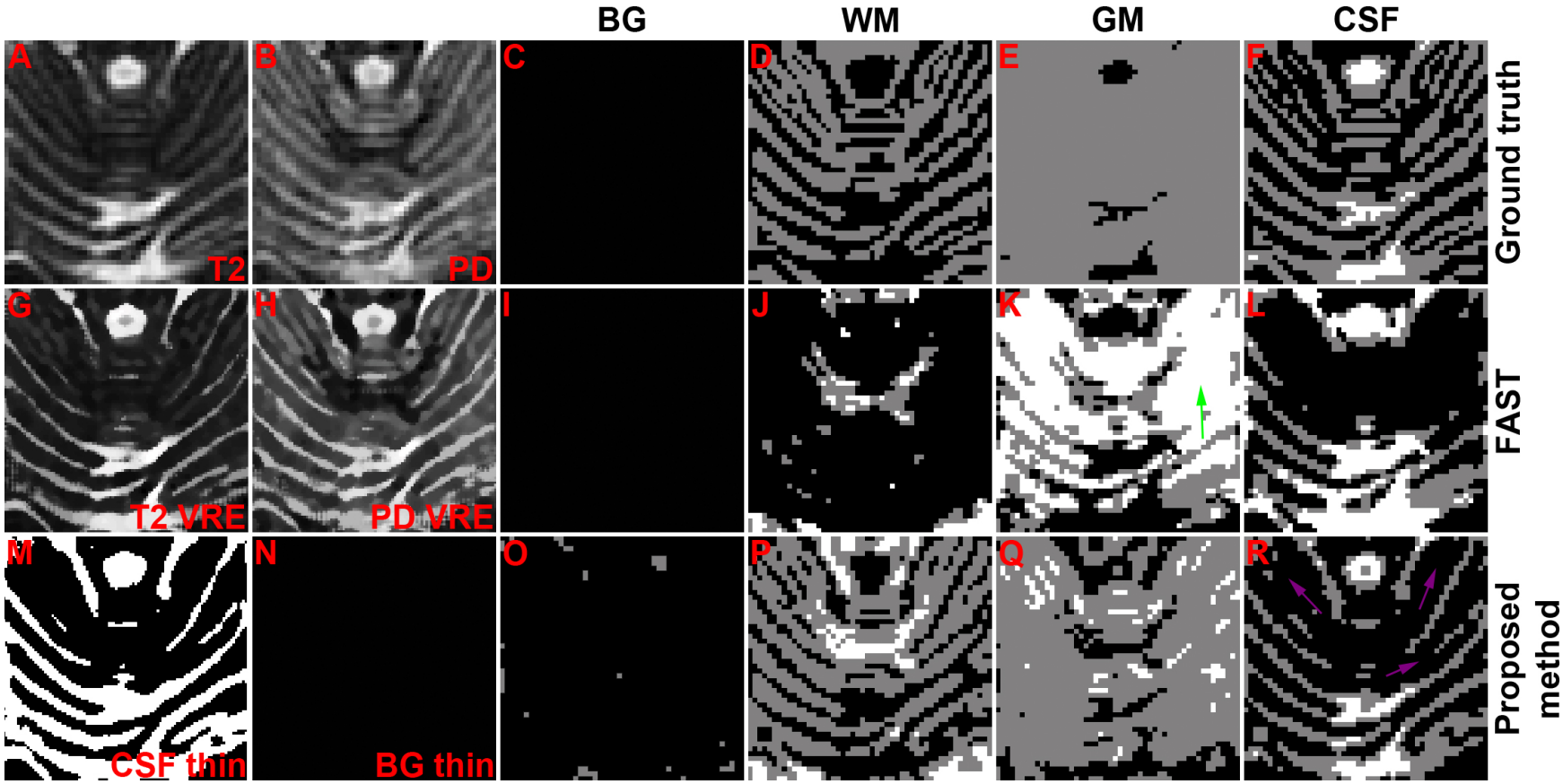

Figure 5.13: **Segmentation of ROI 4**. **A** & **B**) T_2- and PD-weighted data. **C** - **F**) Expert labelling of each tissue class. **G** & **H**) Resolution-enhanced T_2- and PD-weighted data. **I** - **L**) Labelling of each tissue class produced by FAST. **M** & **N**) Binary thin CSF and BG structure maps. **O** - **R**) Labelling of each tissue class produced by the proposed methods. Coloured arrows are referenced in Section 5.4.6.

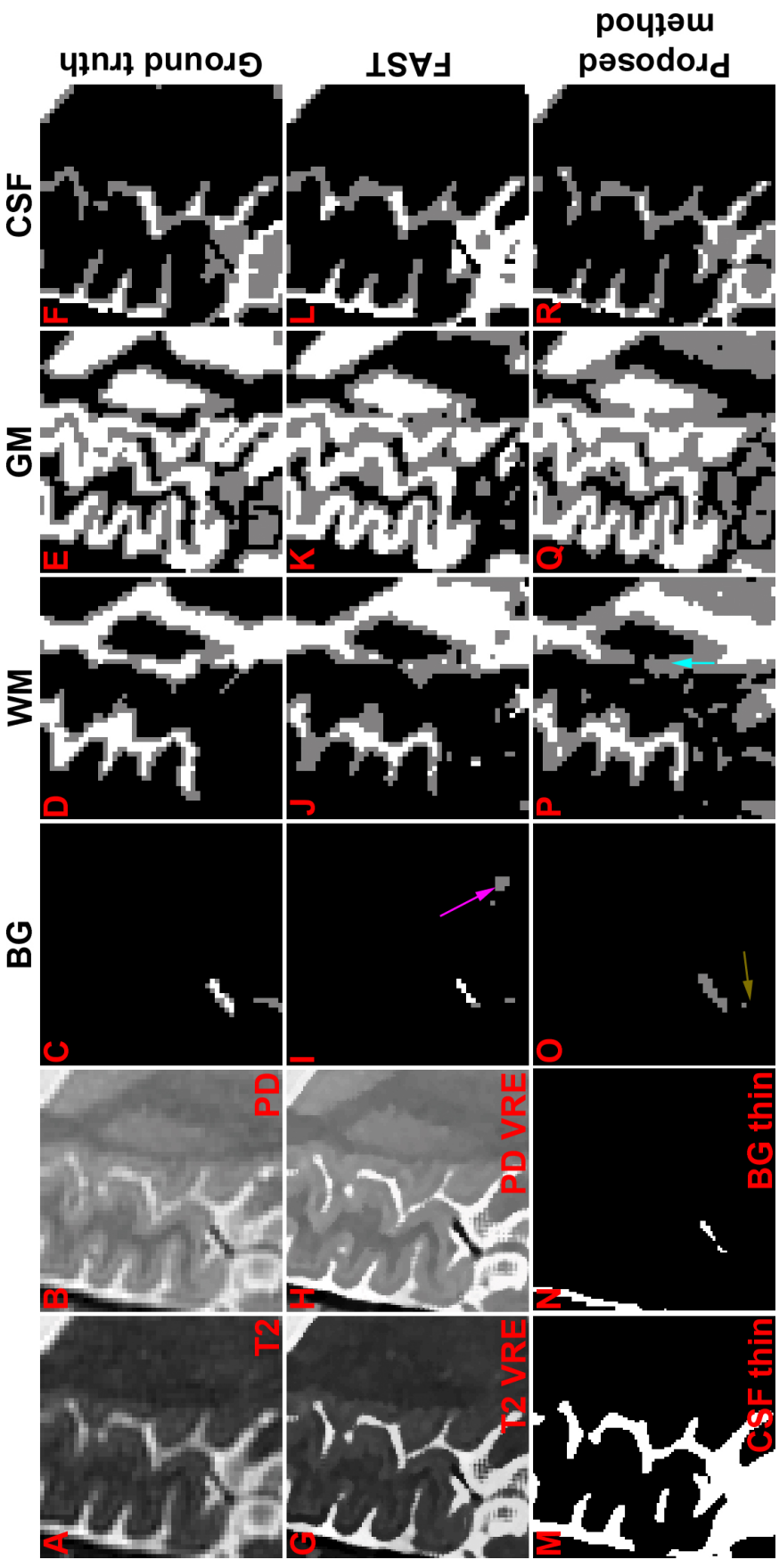

Figure 5.14: **Segmentation of ROI 5**. **A** & **B**) T_2- and PD-weighted data. **C** - **F**) Expert labelling of each tissue class. **G** & **H**) Resolution-enhanced T_2- and PD-weighted data. **I** - **L**) Labelling of each tissue class produced by FAST. **M** & **N**) Binary thin CSF and BG structure maps. **O** - **R**) Labelling of each tissue class produced by the proposed methods. Coloured arrows are referenced in Section 5.4.6.

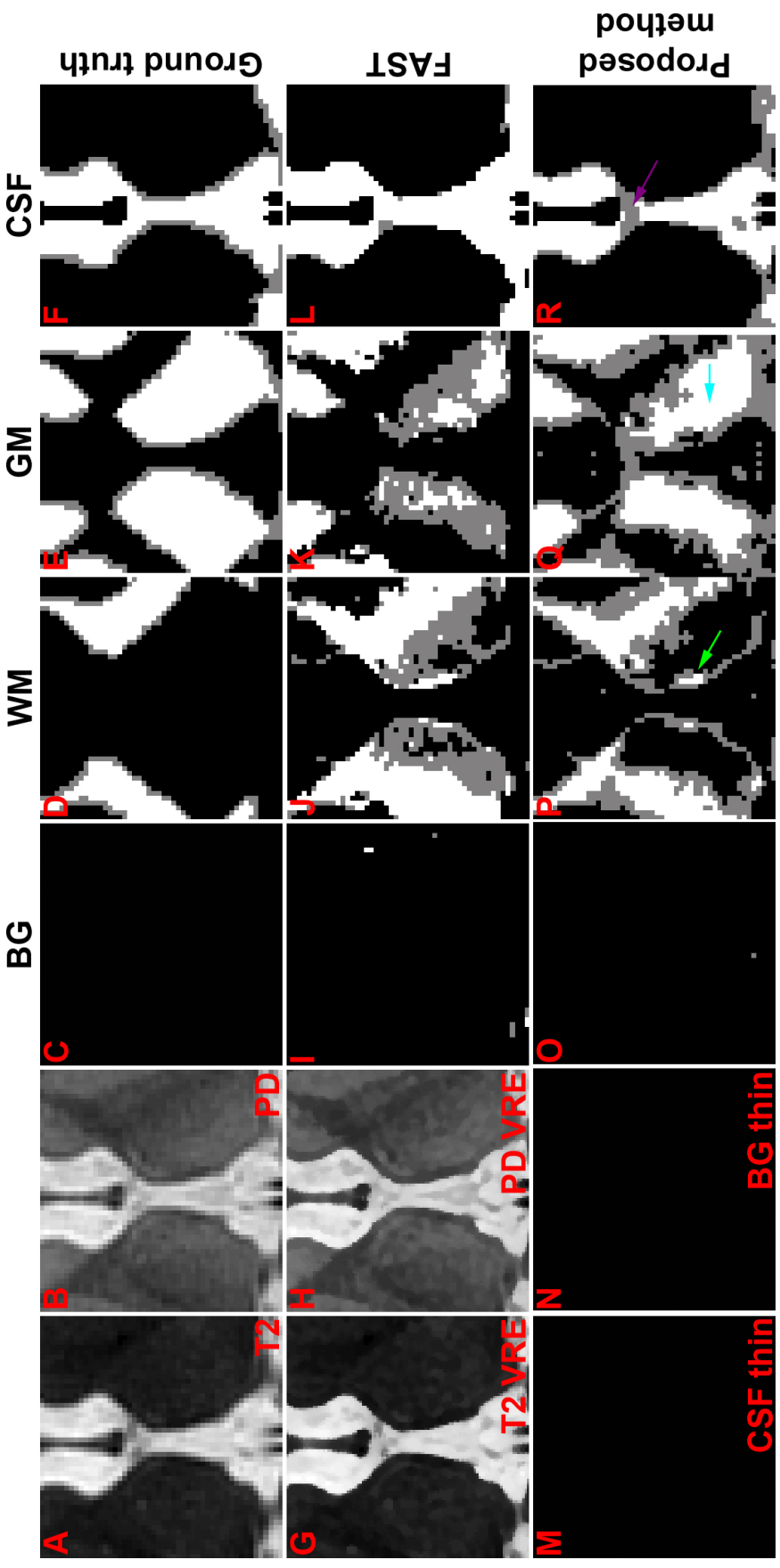

Figure 5.15: **Segmentation of ROI 6**. **A** & **B**) T_2- and PD-weighted data. **C** - **F**) Expert labelling of each tissue class. **G** & **H**) Resolution-enhanced T_2- and PD-weighted data. **I** - **L**) Labelling of each tissue class produced by FAST. **M** & **N**) Binary thin CSF and BG structure maps. **O** - **R**) Labelling of each tissue class produced by the proposed methods. Coloured arrows are referenced in Section 5.4.6.

with an green arrow) is corrected mainly via restoration of the sulcal intensities, observed in panels G and H. Between the sulci, the two WM/GM labels available in the label set are able to characterise the mixing of tissues in the cerebellar cortical voxels well. Substantial improvements also arise from fuller segmentation of deep GM nuclei in ROIs 6 (Fig 5.15) and 7: the edges of the thalamus (aqua arrow) better agree with the expert labelling, and its core contains more pure tissue.

Some erroneous GM labelling was caused by issues in the volume resolution enhancement process. The border of the corpus callosum WM structure with the lateral ventricle in ROI 1 (Fig 5.10, aqua arrow) appears to be well separated into WM and CSF sub-voxels (panels G and H), but on adjacent slices the processing has not worked quite as well causing partial volumes of GM to be labelled. ROIs 2 (Fig 5.11) and 10 both contained occipital regions, and similar to the renderings in Fig 5.8 the poor WM/GM contrast and thinness of the cortex contributed to poor estimation of the cortical boundary. Quite serious overestimation of WM occurred as a result (aqua arrows), with an abundance of WM/GM labels homogenised and even eroded slightly in the presence of nearby large WM regions by the label optimisation. Another source of GM labelling error was in thin WM regions which were incorrectly homogenised into the nearby deep GM nuclei and the convoluted cortex. Examples shown in ROIs 3 and 5 (Figs 5.12 and 5.14, aqua arrows) display poor restoration of the pure WM intensities by resolution enhancement, due to the poor WM/GM boundary contrast and the closeness of the boundaries between the structures.

The accuracy of WM labelling is closely tied to GM since it is its most common neighbouring tissue type and also because the classes' intensity means are not particularly well separated (particularly in T_2) leading to confusion between them. Therefore the majority of the criticisms stated in the previous paragraphs about the quality of GM labelling also apply to WM, and naturally the biggest improvements came again in ROIs 4 (Fig 5.13), 6 (Fig 5.15) and 7, with poorer performance in ROIs 2 (Fig 5.11) and 10. "Ghosts" (first discussed in Section 3.6) bear special mention since thin, dark GM regions adjacent to CSF region boundaries were erroneously assigned partial volumes of WM, as observed in ROI 6 (Fig 5.15, green arrow). Nevertheless, quantitative improvements in WM estimation were made in about half the ROIs, leading to an average 9% reduction in absolute error.

Every ROI except ROI 6 (Fig 5.15) showed substantial quantitative improvement of CSF labelling through both metrics, with an average 31% reduction in absolute error. ROI 6 did not contain any sulci but there was underestimation of the lateral ventricle

(purple arrow) compared to FAST, which was a slight overestimator. The CSF thin structure intensity restoration process and adaptive neighbourhood weighting was very effective in ROI 3 (Fig 5.12) where the presence of many closely-packed sulci led to label homogenisation and in turn massive overestimation of CSF using FAST (blue arrow), and also in ROI 4 (Fig 5.13) where FAST completely mislabelled the rapidly-varying tissue content in the cerebellum (green arrow) through its homogenising and greedy label optimisation. However, not all the sulci in this region were restored by resolution enhancement or labelled correctly (purple arrows) due to their low contrast.

Regions exhibiting failures of correct CSF labelling include small bright blobs evident in ROIs 1 and 3 (Figs 5.10 and 5.12, orange arrows) which were largely omitted from the labelling of both segmentation methods, though FAST was slightly more sensitive. These structures were likely pathological oedema, and since thin structure detection correctly did not flag these blobs (as neither plate nor tube shapes) then resolution enhancement did not restore pure CSF intensities but instead may have smoothed them. Another failure common to most ROIs was slight underestimation of thin cortical sulci which was most apparent in ROIs 2 (Fig 5.11) and 10, likely stemming from the downsampling method employed to enable comparison with the expert-labelled results. In panel M of Fig 5.11 the sulci were actually overestimated (green arrows) by thin structure detection due to their soft edges, and the resolution enhanced data in panels G and H clearly displayed pure CSF intensities for them, but following downsampling of the labelling the CSF content (panel R) seemed much reduced.

The accuracy of estimating BG content was affected by putative expert labelling errors in ROIs 2 (Fig 5.11) and 10 which displayed regions of occipital cortex. Cavities around the falx cerebri (blue arrow) were labelled as WM by Dr Farrall but they were flagged as BG by both segmentation methods, with the labelling of the proposed methods being slightly fuller. FAST's labelling was inferior in a majority of the remaining ROIs and the proposed methods produced an average 36% reduction in absolute error. The improvements in labelling BG mainly stemmed from the exclusion of BG/WM and BG/GM labels from consideration, resulting in reduced noisiness in labelling tissue which appeared slightly dark (apparent in ROIs 1 and 5 (Figs 5.10 and 5.14, pink arrows). Restoration of the pure intensities at the core of vessels generally only led to minor differences in the labelling accuracy, despite the accuracy of the BG localisation in ROI 1 (Fig 5.10, blue arrow). Some vessels were not flagged in B^{BG} leading to mislabelling in ROIs 3 and 5 (Figs 5.12 and 5.14, brown arrows).

Over all the classes combined, the proposed methods were more accurate – using

both quantitative metrics – in every ROI except 2 (Fig 5.11) and 10, where issues with WM/GM contrast and expert labelling errors have been noted. However, the difference in absolute error in these exceptional ROIs was small, at 39 and 33 absolute error units respectively. In total, for all ROIs there was a 21% reduction in absolute error from FAST's results.

5.4.7 BrainWeb dataset

The BrainWeb phantom (described in Section 1.1.6) has commonly been used to test segmentation algorithms (investigated in Section 5.1) as a benchmark dataset, and some of their accuracy statistics are listed in Table 5.1. However, since some segmentation algorithms were not designed to process T_2-PD data and because the statistics were generated from entire volumes rather than small ROIs, then it may not be appropriate to compare these performance measures directly with those listed in this section; furthermore, different pre-processing is likely to have been used. However, the proposed methods were compared against FAST in order to provide a benchmark. FAST did not converge properly when run with dual-channel T_2-PD BrainWeb data despite trying a range of parameter values and even specifying the class means[12] as input; two different WM classes were formed, and GM and CSF content combined in another (likely due to their poor contrast in PD). FAST did perform single-channel segmentation successfully once the data's weighting factor had been specified and since the results for the PD-weighted data were the best of the two, by both absolute error and κ-statistic score measures, then these were tabulated.

This section initially concentrates on evaluating the accuracy of fractional tissue content estimation of the proposed methods, since the exact fractions of classes generating the phantom are known. The results must be interpreted in the context of ideal acquisition and brain structure conditions which cannot be realised in clinical practice: a simplistic model of MRI acquisition is employed, for which anatomical structures generally are not less than several voxels in thickness with perfectly homogeneous content. No bias field effects were employed but the data was generated with several different noise levels (up to 1%, 3%, 5% and 9% of the maximum intensity of each volume) to assess the methods' robustness. BG is not modelled by the phantom within the parenchyma and so any labelling of this class is erroneous.

The quantitative results (listed in Table 5.5) show a steady increase in the abso-

[12] BG = 1, WM = 180, GM = 210 and CSF = 240 for the PD-weighted data. BG = 1, WM = 70, GM = 100 and CSF = 240 for the T_2-weighted data.

FAST (PD only)								
	Absolute error				*RMS error*			
Region	**WM**	**GM**	**CSF**	**All**	**WM**	**GM**	**CSF**	**All**
BW1 (1%)	1.5%	5.1%	4.0%	10.6%	0.094	0.188	0.183	0.279
BW1 (3%)	2.4%	6.5%	4.5%	13.6%	0.134	0.229	0.204	0.335
BW1 (5%)	3.4%	9.5%	6.5%	19.4%	0.177	0.305	0.263	0.440
BW1 (9%)	7.2%	17.6%	12.9%	32.7%	0.324	0.492	0.423	0.725
BW2 (1%)	2.0%	3.8%	1.8%	7.6%	0.091	0.152	0.113	0.210
BW2 (3%)	3.2%	5.5%	2.2%	11.0%	0.140	0.202	0.136	0.281
BW2 (5%)	4.2%	7.8%	3.8%	15.9%	0.182	0.268	0.202	0.382
BW2 (9%)	8.5%	16.7%	9.8%	35.2%	0.339	0.495	0.398	0.720
Proposed method								
BW1 (1%)	1.2%	2.5%	1.4%	5.1%	0.059	0.090	0.068	0.128
BW1 (3%)	1.8%	3.0%	1.3%	6.1%	0.087	0.108	0.063	0.152
BW1 (5%)	2.4%	3.8%	1.4%	7.7%	0.114	0.134	0.070	0.189
BW1 (9%)	3.8%	5.5%	1.8%	11.2%	0.181	0.202	0.087	0.285
BW2 (1%)	1.4%	1.9%	0.5%	3.8%	0.061	0.067	0.034	0.099
BW2 (3%)	2.0%	2.6%	0.6%	5.2%	0.087	0.095	0.038	0.135
BW2 (5%)	3.0%	3.6%	0.7%	7.3%	0.122	0.130	0.045	0.185
BW2 (9%)	4.8%	5.7%	1.0%	11.5%	0.189	0.200	0.067	0.283

Table 5.5: **Error statistics for the proposed segmentation method (using both T_2- and PD-weighted data) and FAST (using T_2-weighted data only) over two ROIs from the BrainWeb dataset**. Errors are listed to 3 decimal places for the RMS error, and to 1 decimal place for the absolute error percentages. **BW1**: BrainWeb volume, region 1 (as pictured in Fig 5.16). **BW2**: BrainWeb volume, region 2 (as pictured in Fig 5.17). Noise level of the BrainWeb volume is noted in brackets. 6400 voxels were present in each ROI. The smallest absolute and RMS errors between the proposed method and FAST are highlighted in red.

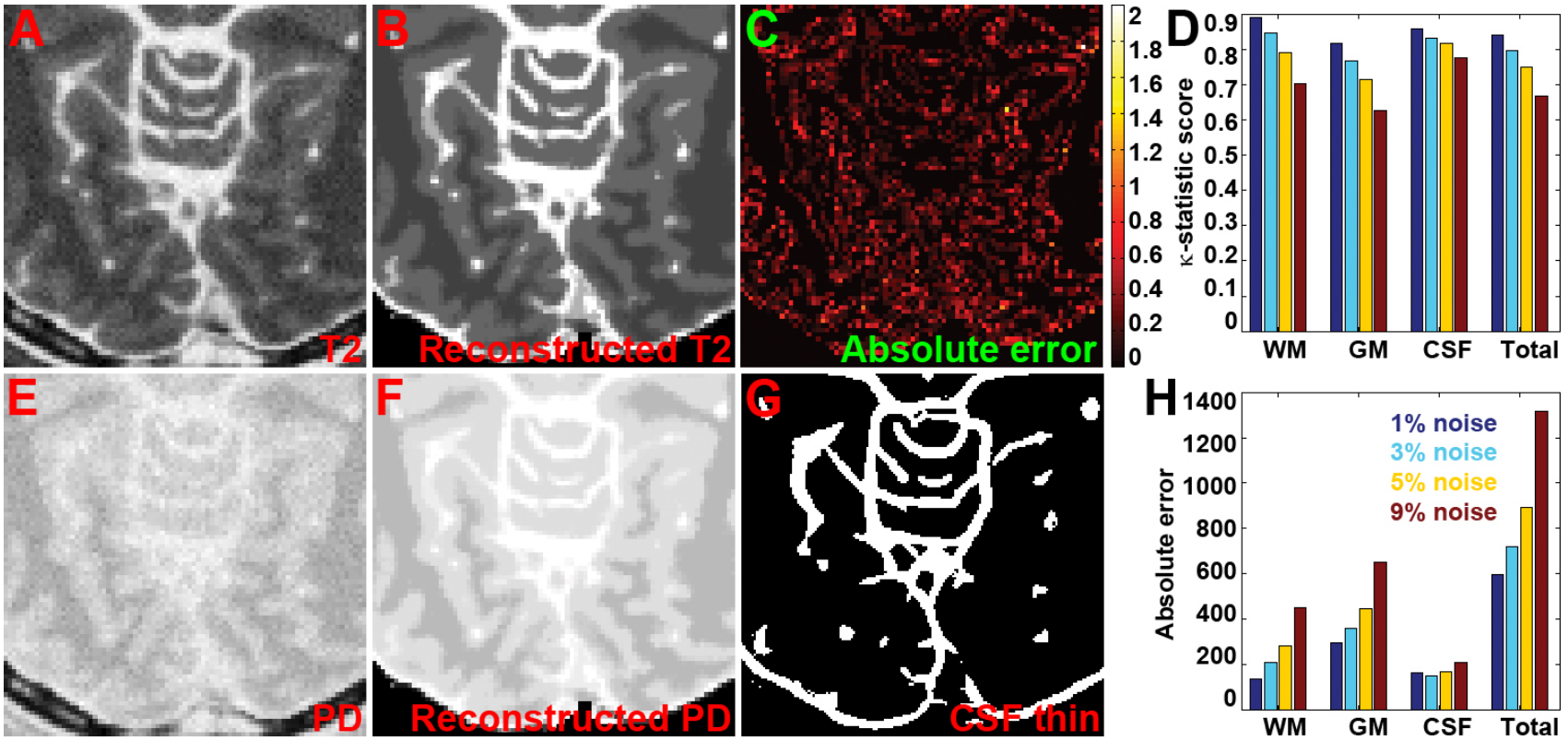

Figure 5.16: **Segmentation of a region of occipital cortex from the BrainWeb volume using the proposed segmentation methods**. **A** & **E**) T_2- and PD-weighted data from the 3% noise volume. **B** & **F**) T_2- and PD-weighted data reconstructed from the labelling and class means of the 3% noise volume, using the mixel model. **C**) Absolute total labelling error for all classes from the 3% noise volume. **D**) κ-statistic scores for this ROI. **G**) CSF thin structure map for the 3% noise volume. **H**) Absolute error values for this ROI.

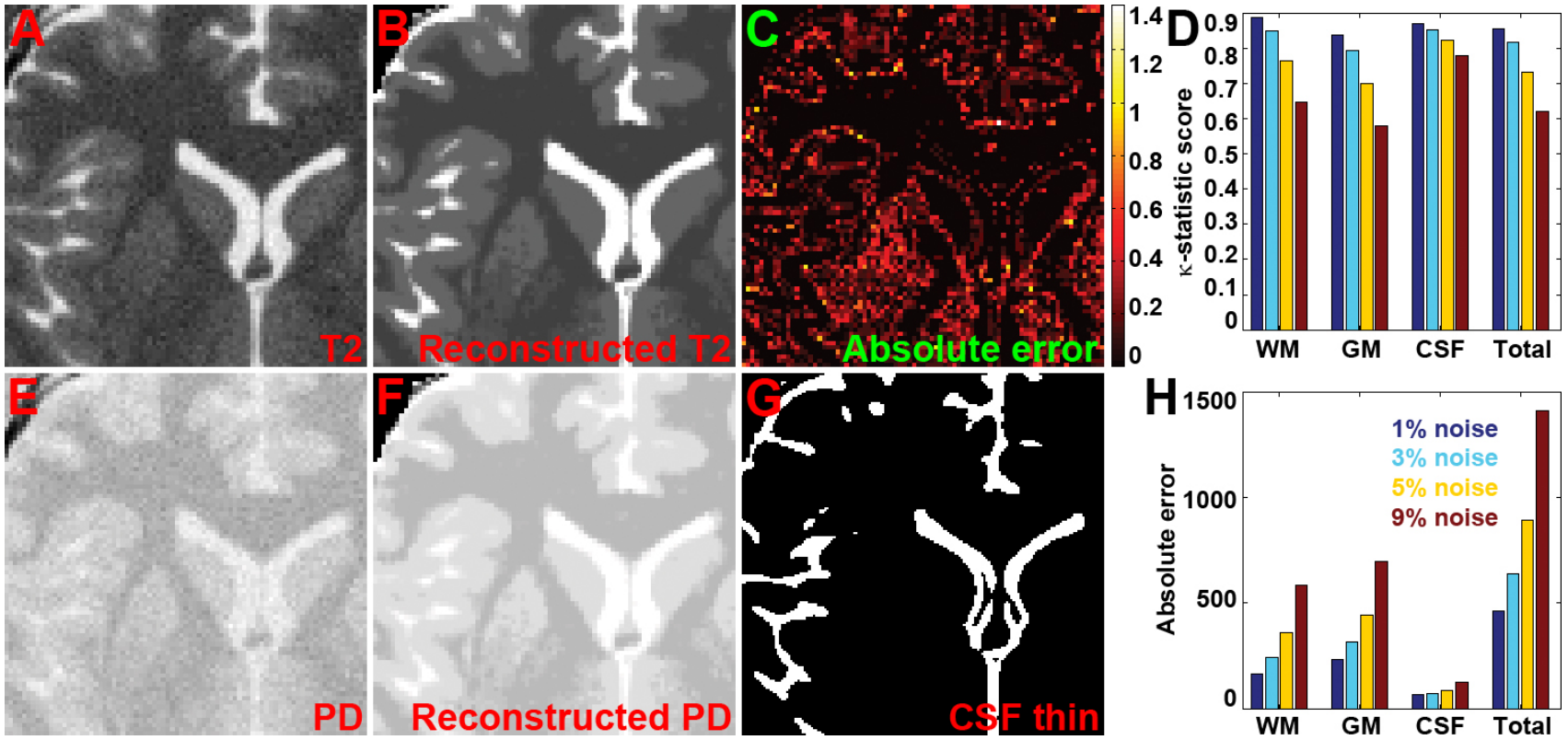

Figure 5.17: **Segmentation of a frontal region from the BrainWeb volume using the proposed segmentation methods**. **A** & **E**) T_2- and PD-weighted data from the 3% noise volume. **B** & **F**) T_2- and PD-weighted data reconstructed from the labelling and class means of the 3% noise volume, using the mixel model. **C**) Absolute total labelling error for all classes from the 3% noise volume. **D**) κ-statistic scores for this ROI. **G**) CSF thin structure map for the 3% noise volume. **H**) Absolute error values for this ROI.

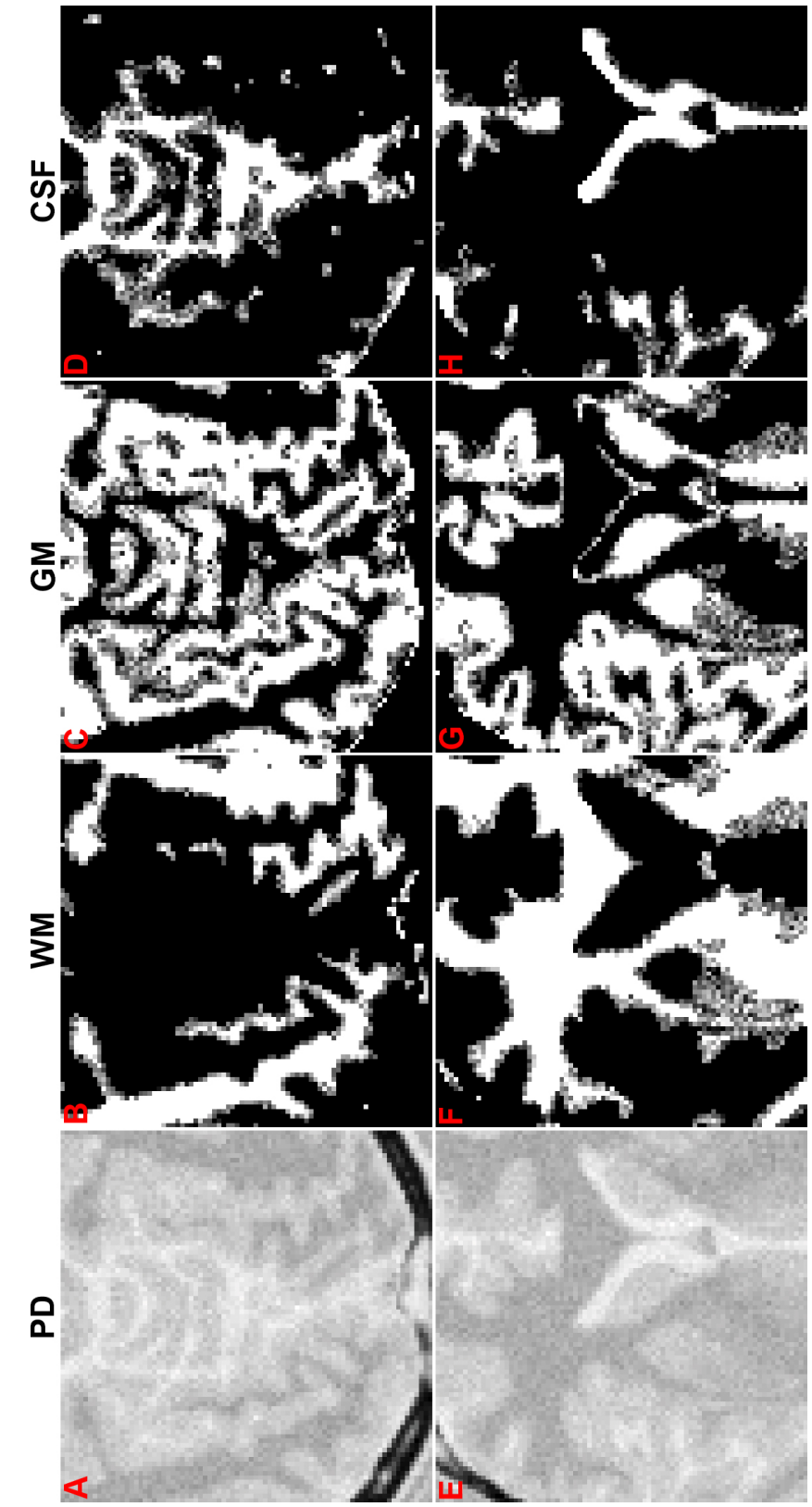

Figure 5.18: **Soft segmentation results produced by FAST [Zhang et al., 2001] for two ROIs from the BrainWeb phantom, using PD-weighted data only**. **A** & **E**) PD-weighted data from the ROIs. **B-D**) Soft segmentation results for the first ROI. **F-H**) Soft segmentation results for the second ROI.

lute and RMS errors with increasing noise levels for both ROIs studied, with the pace of these changes varying from almost linear to slightly exponential in the noise level. GM consistently had the greatest errors of any tissue class, likely since it has the most complex boundaries, with WM the next worst as GM's most common neighbour. The magnitude of the signed volume errors (data not shown) compared to the absolute errors was fairly small for the proposed method, indicating that discrepancies largely came from slight over- or under-estimation of content at tissue boundaries without significant trends. This statement was confirmed by the localisation of absolute errors summed over all classes to boundaries in panel C of Figs 5.16 and 5.17. FAST's segmentation results were worse (by both the absolute and RMS error measures) for each tissue class and showed considerable degradation at high noise levels; the advantages of using dual-channel data are particularly apparent since it improved class definition in the intensity space, and also because the noise (visible as very rough structure edges in Fig 5.18) was smoothed by the resolution enhancement process in a manner consistent between the image weighting factors. Furthermore, correct labelling of the sulci was maintained by the robust thin CSF structure prior – and since FAST did not employ this kind of prior then there was considerable confusion between the GM and CSF classes.

Reconstructions (panels B and F) of the intensity data from the proposed methods' output – as class content fractions and the classes' means – produced images which were very similar to noise-free acquisitions from the phantom. However, a few patches of noise were not removed and the gradiated content of the deep GM nuclei is not as fully nor as smoothly apparent as in the ground truth, likely because of the small range of WM/GM labels made available.

The proposed method's abilities of working with entire volumes of $2 \times 2 \times 2$ downsampled low-resolution data were also tested, with quantitative results at the upscaled level listed in Table 5.6 and the segmentation results from the same region as pictured in Fig 5.17 plotted in Fig 5.19. Compared to the results tabulated in Table 5.5 the accuracy at low noise levels is considerably lower, but the drop-off of performance at higher noise levels is much improved. Since the downsampling affects high-frequency structures in the volume the most – notably noise, complicated WM/GM boundaries (marked with a blue arrow in Fig 5.19) and the thin cortical sulci (orange arrow) – then the poorer sensitivity to these structures as well as noise is expected. The only parameter changed when processing these volumes was halving the expected scale of thin structures for the thin structure detection process, and further parameter tuning

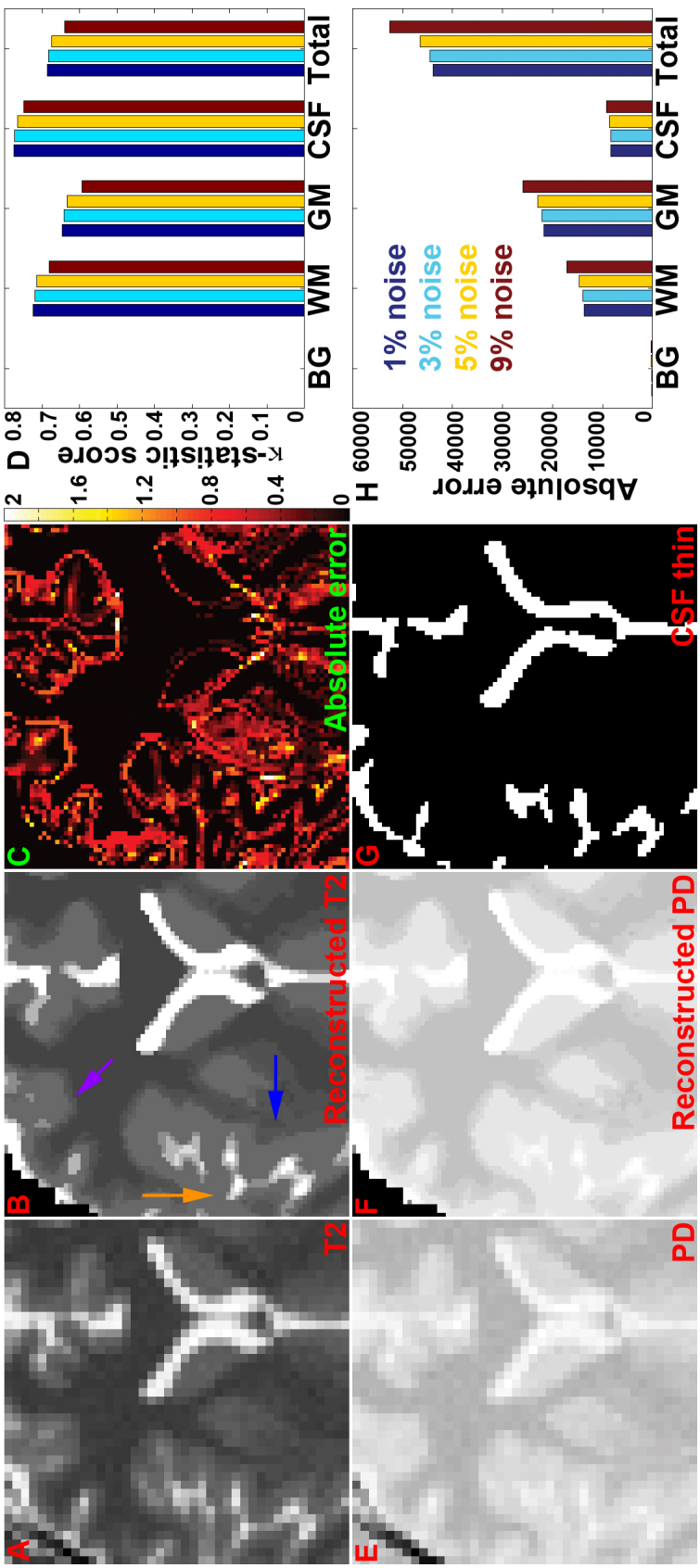

Figure 5.19: **Segmentation of a frontal region from the $2 \times 2 \times 2$ downsampled BrainWeb volume (3% noise data shown) using the proposed segmentation method**. **A** & **E**) T_2- and PD-weighted data (the same data pictured in Fig 5.17 is shown). **B** & **F**) Reconstructions of intensity from the labelling and class means, using the mixel model. **C**) Absolute total labelling error for all classes. **D**) κ-statistic scores for the entire volume at a range of noise levels. **G**) CSF thin structure map. **H**) Absolute error values for the entire volume at a range of noise levels. 1,970,280 brain voxels were present in the volume.

	Absolute error				*RMS error*			
Region	**WM**	**GM**	**CSF**	**All**	**WM**	**GM**	**CSF**	**All**
BWDSV (1%)	3.8%	6.0%	2.3%	12.0%	0.173	0.215	0.130	0.305
BWDSV (3%)	3.8%	6.1%	2.3%	12.1%	0.175	0.216	0.130	0.307
BWDSV (5%)	4.0%	6.3%	2.4%	12.7%	0.183	0.225	0.133	0.319
BWDSV (9%)	4.7%	7.1%	2.5%	14.3%	0.206	0.247	0.141	0.351

Table 5.6: **Error statistics for the proposed segmentation method over the entire BrainWeb volume, downsampled by ratio** $2 \times 2 \times 2$. Errors are listed to 3 decimal places for the RMS error, and to 1 decimal place for the absolute error percentages. **BWDSV**: Downsampled BrainWeb volume. Noise level of the BrainWeb volume is noted in brackets. 1,970,280 brain voxels were present in the original volume.

may have lessened unwanted effects such as the thickening of WM/GM boundaries (purple arrow). While the two ROIs studied previously contained a disproportionate amount of complicated anatomical structure, the entire volume contained many less complicated regions, contributing to the fairly low absolute and RMS errors obtained when processing the entire volume.

5.4.8 Comparison of performance with the unified segmentation method of van Leemput *et al.*

Since the work in this thesis was inspired by the work presented in [van Leemput et al., 2003], it was natural to compare some of its segmentation results on the LBC1936 dataset with those of the proposed methods. Quantitative evaluation of the results in Table 5.7 revealed much improved estimation of GM and CSF content using the proposed methods except in ROI 6 (previously pictured in Fig 5.15), where (as noted in Section 5.4.6) there was underestimation of the lateral ventricle. WM content was better estimated by the unified segmentation method in ROIs 2, 9 and 10 – though not by a substantial amount – and the former and latter cases can be explained by the putative expert labelling errors between WM and BG around the falx cerebri noted in Section 5.4.6. In ROI 9 the slight differences were apparent at the WM/GM boundaries, which tended to agree more with the expert. BG estimation was better in most cases (except ROIs 3, 5 and 6) in the unified segmentation method's results, but the discrepancy in each case can be fully explained. Ignoring ROIs 2 and 10 which contained putative ex-

pert labelling errors as well as ROIs 1, 6 and 8 which contained only minor differences around a vessel and the choroid plexus, the only ROIs with more significant differences were 3 and 4. The BG in ROI 3 (pictured in Fig 5.12) was underestimated by the unified segmentation method, whereas the proposed methods attempted to label more of the vessel content with partial volume labels – some of which was not labelled as such by the expert. In ROI 4 (pictured in Fig 5.13) the proposed methods erroneously labelled partial volumes of BG content due to the presence dark WM intensities in the PD-weighted data on adjacent slices following resolution enhancement. The BG labelling in this case could likely be rectified with more robust BG class parameter estimation, which produced high mean BG parameters.

Overall, there was a 40% decrease in absolute error for all classes in all the ROIs. A direct comparison between the methods over ROI 5 is presented in Fig 5.20, with the most striking change being the improved estimation of pure GM content in the tightly-folded temporal cortex (marked with a blue arrow), GM in the putamen (green arrow) and GM in the thalamus (orange arrow). The estimation of cortical sulci in the temporal cortex is also much better (purple arrow). Many of these differences can be attributed to the inability of the unified segmentation method to express partial volume labelling properly at the sub-voxel level, directly affecting correct segmentation of the thalamus and thin cortical sulci as well as indirectly causing poor estimation of the WM and GM class parameters. Thin structure detection has aided the fuller labelling of the blood vessel and the thin sulci, and resolution enhancement has improved discrimination of the WM/GM boundary in the temporal cortex.

5.4.9 Robustness to pathological data

A version of the BrainWeb dataset is available with multiple sclerosis lesions added as a fifth main tissue class, and the segmentation results using the proposed methods are presented in Fig 5.21. The multiple sclerosis lesions in the second ROI are largely indistinguishable from CSF content so the CSF class picks up both. Crucially, the segmentation of the four main tissue classes is largely unaffected, although the partial volume GM content in the thalamus becomes a little weak (marked with a purple arrow). This was possibly caused by slight complications with estimating the GM class parameters, due to rings of GM (orange arrow) forming around the multiple sclerosis lesions. Bands of GM-like intensity were formed in the resolution-enhanced data due to the lesions' diffuse boundaries; as noted in the Discussion of Chapter 4, these re-

Unified segmentation method [van Leemput et al., 2003]														
	Absolute error					*κ-statistic score*					*Volume error*			
ROI	**BG**	**WM**	**GM**	**CSF**	**All**	**BG**	**WM**	**GM**	**CSF**	**All**	**BG**	**WM**	**GM**	**CSF**
1	3.5	621	664.5	137	1426	0.630	0.427	0.193	0.723	0.409	-1.5	577	-447.5	-128
2	7	541.5	676.5	269	1494	0.000	0.370	0.191	0.474	0.256	7	442.5	-390.5	-59
3	40.5	454.5	686.5	333.5	1515	0.051	0.227	0.250	0.471	0.268	-24.5	-169.5	-19.5	213.5
4	0.5	670	1082.5	589	2342	0.000	0.069	0.037	0.241	0.066	0.5	-272	439.5	-168
5	19	702.5	816	149.5	1687	0.301	0.366	0.287	0.662	0.348	-8	683.5	-586	-89.5
6	2	1194	1194.5	69.5	2460	0.000	0.174	0.039	0.883	0.295	2	1190	-1156.5	-35.5
7	5	310	359	86.5	759	0.499	0.663	0.525	0.816	0.654	-1.5	-243	290	-45.5
8	1.5	722.5	755.5	49	1528	0.666	0.215	0.270	0.756	0.295	-1	383.5	-377.5	-5
9	7.5	153.5	246	102.5	507	0.695	0.781	0.669	0.769	0.713	-1	-53.5	-8	62.5
10	33.5	461	635.5	257	1387	0.357	0.458	0.195	0.356	0.261	-15.5	421	-247.5	-158
Proposed method														
1	6.5	225.5	295	114	641	0.433	0.716	0.607	0.773	0.671	0.5	-17.5	85	-68
2	47	566.5	467.5	201	1282	0.000	0.313	0.359	0.574	0.325	47	494.5	-369.5	-172
3	32	430	540.5	195.5	1198	0.506	0.234	0.284	0.642	0.351	3	-116	133.5	-20.5
4	20.5	372.5	250	286	929	0.000	0.472	0.224	0.577	0.431	20.5	296.5	-113	-204
5	15	319	381.5	101.5	817	0.370	0.554	0.542	0.785	0.579	-10	117	-74.5	-32.5
6	0.5	453.5	488	132	1074	0.000	0.387	0.492	0.774	0.515	0.5	257.5	-135	-123
7	7	254.5	287	84.5	633	0.515	0.662	0.531	0.819	0.645	7	-132.5	200	-74.5
8	1.5	463	476	44.5	985	0.666	0.326	0.410	0.767	0.413	1.5	88	-52	-37.5
9	14.5	176	241	88.5	520	0.329	0.759	0.681	0.780	0.704	5.5	0	64	-69.5
10	69.5	479	445.5	134	1128	0.373	0.394	0.334	0.679	0.371	69.5	387	-339.5	-117

Table 5.7: **Error statistics for the proposed segmentation method and the unified segmentation method [van Leemput et al., 2003] over expert-labelled ROIs from the LBC1936 dataset**. Errors are listed to 3 decimal places. The smallest absolute error and largest κ statistics between the proposed method and the unified segmentation method are highlighted in red; equal values are highlighted in blue. 2500 voxels were present in each ROI.

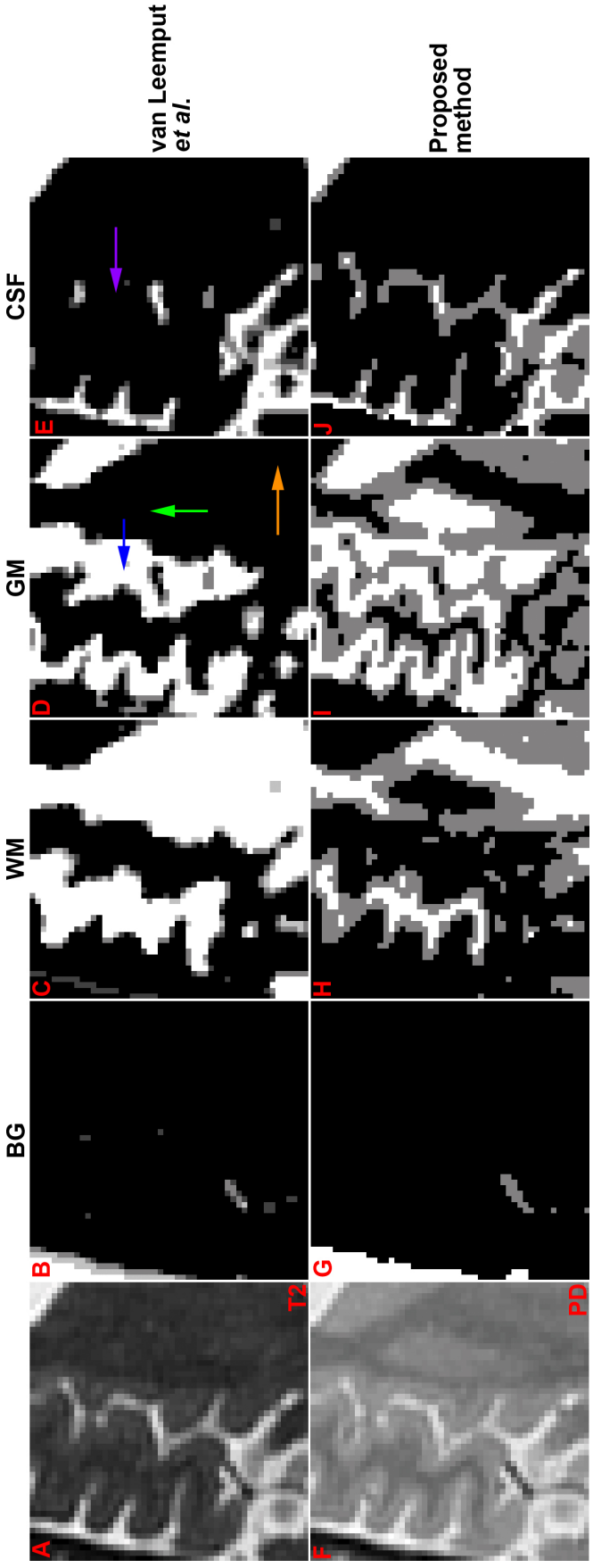

Figure 5.20: **Comparison of the soft segmentation results for ROI 5 produced by the method described in [van Leemput et al., 2003] with the methods described in this thesis**. **A** & **F**) T_2- and PD-weighted data. **B-E**) Soft segmentation results produced by using the method described in [van Leemput et al., 2003]. **G-J**) Soft segmentation results produced using the methods described in this thesis. Coloured arrows are referenced in Section 5.4.8.

gions should be detected and removed from normal resolution enhancement processing if possible.

Atrophy is common in aged brains and an example from the LBC1936 dataset is shown in Fig 5.22. The labelling produced by the proposed methods overestimates WM content, but it does not form a degenerate result. While the bright intensity contrast and strong thin structure prior map ensures CSF content is estimated well, poor GM contrast – particularly in the T_2-weighted data – means the segmentation result in posterior cortical regions is poor (marked with blue arrows), there is sometimes confusion with BG (green arrows), and many GM structures such as the caudate and putamen are incorrectly labelled (orange arrows).

5.4.10 Processing efficiency

Excluding the pre-processing, thin structure detection and volume resolution enhancement stages then the segmentation process can be broken into three iterative elements: initialisation via FCM clustering (Section 5.3.2), the R-step (Section 5.3.4), and both the remainder of the E-step (Section 5.3.3) and the M-step (Section 5.3.5). The important requirement was that segmentation of a full volume should complete within a few hours using a modern desktop PC.

The remainder of the E-step and the M-step took the longest at 611.6 ± 38.2 seconds (mean $\pm$ standard deviation over three trials over the volume containing ROI 5 in the LBC1936 dataset) per EM iteration, largely because it was written purely in MATLAB code. Initialisation was the fastest element, taking 89.1 ± 3.3 seconds per FCM iteration as the neighbourhood processing was relatively simple and significant portions of the membership and centroid re-estimation code were implemented in compilable MEX files. The R-step was expensive due to the examination of $(\#M)^2$ label interactions at each neighbour of every voxel, but MEX file implementation meant each sub-iteration only took 137.4 ± 1.6 seconds. Since 4-5 R-step sub-iterations were performed every E-step (depending on the choice of the ambiguity reduction schedule β_z of the neighbourhood context label prior) then both parts took about the same time in each EM iteration.

Factors contributing to changes in processing times for the proposed segmentation methods include the size of the label set M (and in turn M^{frac}, the range of fractional levels allowed) and the physical dimensions of neighbourhood weighting filters (which can be influenced by the volume upscaling ratio L) in addition to the resolution-

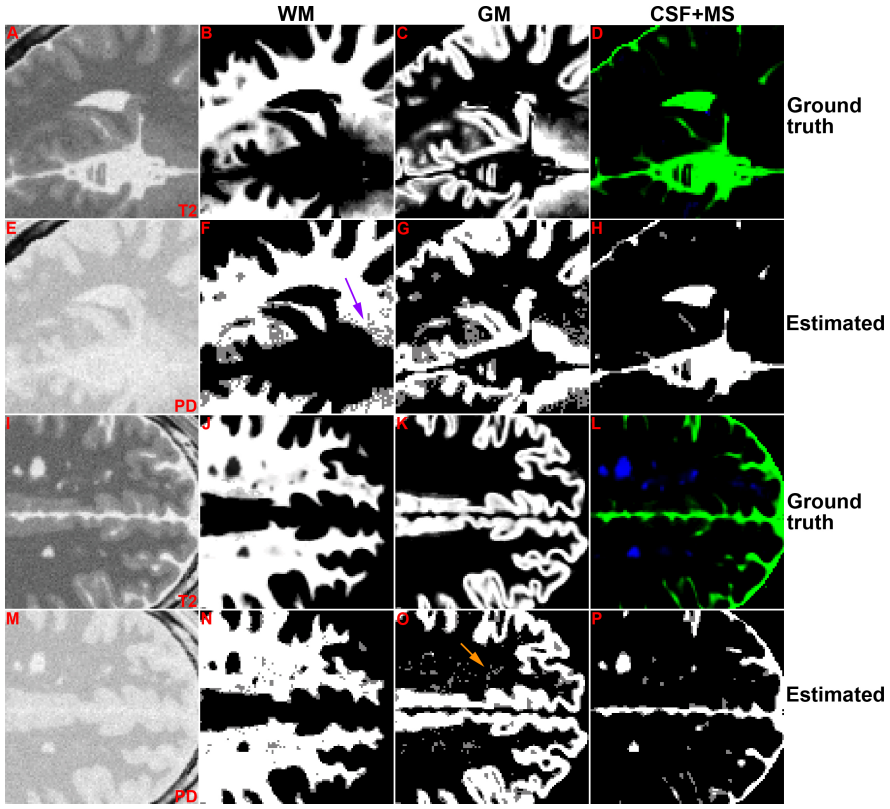

Figure 5.21: **Examination of the robustness of the proposed methods to multiple sclerosis lesions**. **A** & **E** and **I** & **M**) T_2- and PD-weighted data from two ROIs, taken from volumes with 3% added noise. **B**-**D** & **J**-**L**) Ground truth class content proportions, where CSF content is presented in shades of green and MS lesion content in shades of blue. **F**-**H** & **N**-**P**) Estimated class content proportions using the proposed methods.

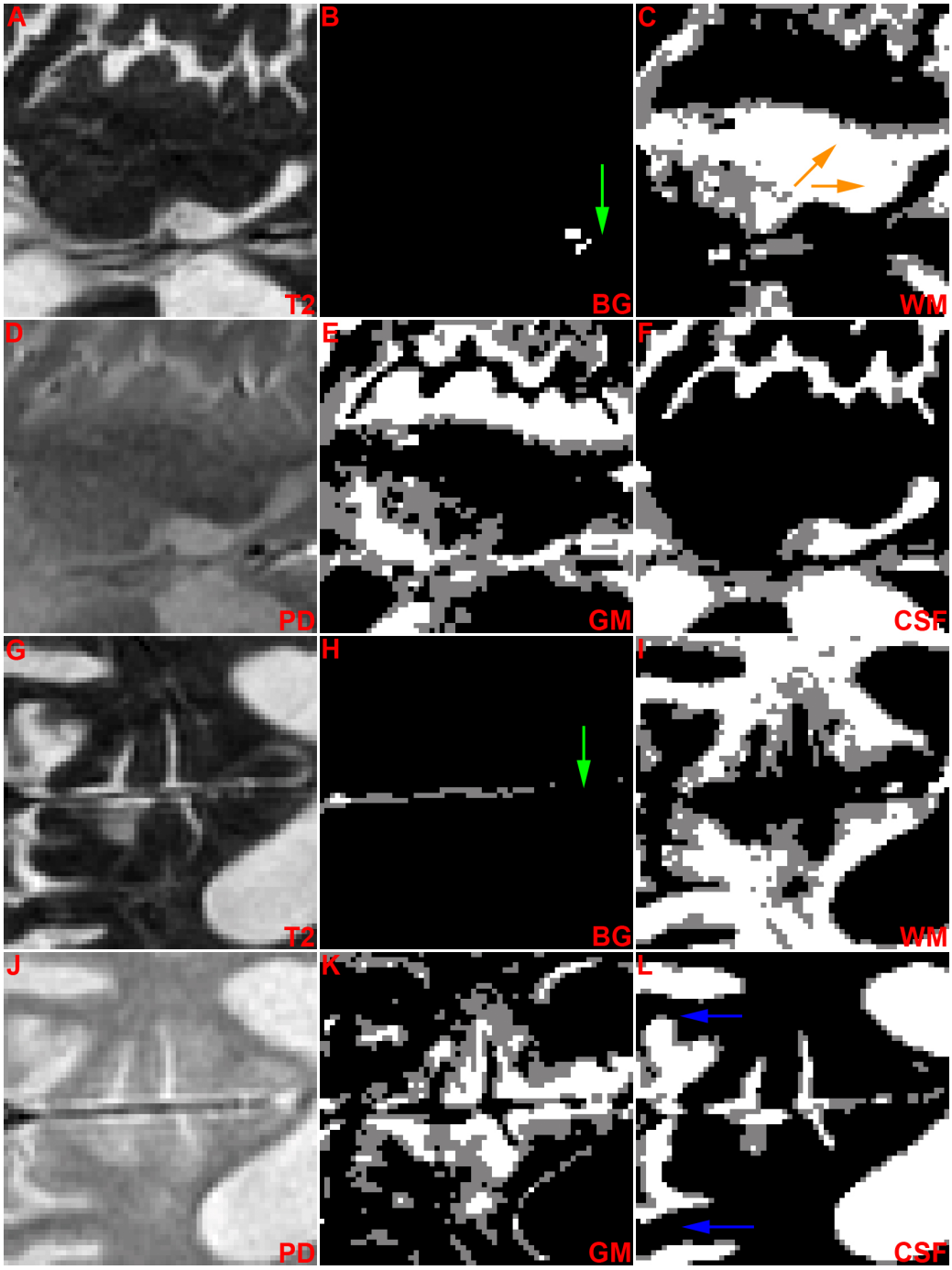

Figure 5.22: **Examination of the robustness of the proposed methods to atrophied brain tissue**. **A** & **D** and **G** & **J**) T_2- and PD-weighted data from medial and posterior ROIs. **B**-**C**, **E**-**F**, **H**-**I** & **K**-**L**) Estimated class content proportions.

enhanced volume size (dependent on the number of voxels and L). Similar to the previous analyses of computational complexity, by assuming most arithmetic operations cost only a multiple of the input data size then the computational complexity overall is linear with the number of voxels.

5.5 Discussion

The entire proposed system, including the segmentation stage, should be evaluated in terms of the original objectives stated in Section 1.4; many have already been achieved in Chapters 3 and 4. Concerning the remaining objectives, Section 5.2 has presented an investigation of data modelling assumptions and testing in Section 5.4 has followed the guidelines expressed in [Udupa et al., 2006]: the demonstrations were performed using expert-labelled data from several different aged individuals (Section 5.4.6), displaying a wide range range of anatomical features, different levels of tissue contrast as well as pathology. In addition, more precise examinations of partial volume labelling accuracy were made using the BrainWeb phantom (Section 5.4.7). The sequential implementation of each of the proposed system's components in Section 5.4 had beneficial effects on the tabulated accuracy metrics (Table 5.3) and in the perceived quality (Fig 5.9), suggesting that these steps may also be beneficial to add (even in isolation) to existing segmentation frameworks. Finally, the segmentation model (detailed in Section 5.3.1) employed both T_2- and PD-weighted data to improve class distinctiveness and increase robustness to noise, and the outputs of previous chapters were put to work in improving data quality as well as providing tissue localisation priors and special neighbourhood weighting filters for thin contrasting structures.

The results produced from the bias-free BrainWeb data – where the ground truth was precisely known allowing comparison with other works in the literature as well as precise examination of the fractional content estimation differences – was affected by the coarse range of WM/GM labels available, which harmed the potential accuracy (expressiveness) in the gradiated content of the deep GM nuclei. However, at structure boundaries where the sub-voxels are expected to have mostly close-to-pure intensities, discretisation errors are mitigated in $\hat{M}$ at the original resolution even using low L and small M^{frac}[13]. Choosing a small M^{frac} also keeps the computational cost of the

[13]For example, a range of sub-voxel label fractions as small as $M^{frac} = \{0, 0.5, 1\}$ alongside $L = [2,2,2]$ are potentially able to discriminate $\left(\left[\text{size}(M^{frac}) - 1\right]\prod L\right)^{-1} = 0.0625$ fractional differences at the voxel level.

R-step in check, which is proportional to the square of the number of labels in M, and this decision ensures that a good decision distance is established from pure tissue labels (otherwise degenerate results may be obtained, as observed in Fig 5.6 using the larger label set). Another way to ensure the correct labelling of pure tissue regions is to introduce a prior on the fractional levels taken for each class [Pham and Prince, 2004], which could pave the way for generative models of tissue mixing.

In general, the errors for two small – though structurally complex – 80×80 voxel regions from the BrainWeb volume were generally small in magnitude, the proposed method outperformed FAST working with single-channel data (Table 5.5), and the intensity reconstructions appeared similar to the 0% noise data (data not shown). In Table 5.5 the proposed method's deviation from the ground truth ranged from 4% to 12% of the possible maximum absolute error depending on the noise level, and outperformed most of the RMS errors on entire volumes obtained from competing methods' publications in Table 5.1. Only three methods were found to deliver similar ([Pham, 2001][Huang et al., 2006]) or slightly better ([Bricq et al., 2008], giving 0.03 less RMS error for WM and GM) accuracy. Of these methods, [Huang et al., 2006] employed active-contour-based segmentation which has been examined in more detail in Section 6.4.1, and [Bricq et al., 2008] used an EM-based approach with the basics similar to that presented in this chapter.

However, the caveats stated in Section 5.4.7 must be bourne in mind when comparing these statistics. The image weighting factors processed were not the same – either only T_1, or T_1 and T_2 – since results incorporating PD and quoting the RMS or absolute error could not be located in the literature, and some volumes had bias field effects added. Furthermore, entire volumes were usually processed by the methods quoted in Table 5.1, which would likely contain a smaller proportion of regions with as complicated boundaries as those observed in the two small ROIs studied. This proposal was verified by the surprisingly good performance of the proposed method (Table 5.6) when processing the downsampled volume; despite the elimination of much fine detail, the error values remained low. A round-up of these kinds of soft segmentation algorithms performed on the same dataset with the same pre-processing methods applied would be worthwhile to quantify the differences and qualify the peculiarities of their labelling in future publications.

Criticism can be levelled at the assumption of 50%/50% content for voxels containing two tissue classes in the analysis of absolute labelling accuracy for the LBC1936 ROIs, since Dr Farrall did not specify the fractional content values. Precise estima-

tion of the fractional class contents through visual analysis alone is likely to introduce additional estimation error as well as considerably increase the time taken to perform manual labelling, neither of which are desirable. Nonetheless, an overall 21% reduction in absolute error was observed in the labelling accuracy (with the statistics listed in Table 5.4) when using the proposed methods compared to FAST [Zhang et al., 2001], and a small increase in the κ-statistic score was present for the vast majority of classes in most ROIs. The main improvements stemmed from better estimation of the extent of thin structures, greater sensitivity to some WM/GM partial volumes through improved GM class definition, and banning the consideration of unlikely class mixtures such as BG/WM to improve WM labelling. These effects in addition to much improved partial volume structure labelling were also apparent in comparisons with the unified segmentation method [van Leemput et al., 2003] to a greater extent, resulting in an overall 40% reduction in absolute error.

In two ROIs the proposed methods performed worse than FAST, and a similar degradation in accuracy was observed in comparisons against the unified segmentation method. BG and WM content estimation in ROIs 2 and 10, which displayed regions of occipital cortex, was hindered by poor GM contrast (illustrated by the reducing contrast levels moving from anterior to posterior regions in Fig 6.1) as well as by the putative mislabelling of a BG structure (marked with a blue arrow in Fig 5.11) by the expert. Some other instances of putative mislabelling of blood vessels were noted in Section 5.4.8. In general, the poor resolution of blood vessels at the original voxel level resulted in sub-optimal detection and subsequent restoration of their pure intensities (as noted in Section 4.4), which then led to confusion in their labelling with similarly-intense WM. Another factor contributing to labelling inaccuracies was an artifact named "ghosting" (first described in Section 3.5.1) which was apparent as dark lamina neighbouring large CSF regions, causing WM to be mistakenly labelled next to the ventricles in Figs 5.10 and 5.15. No practical work-around could be found during segmentation stage since WM/CSF partial volume voxels did exist elsewhere, but correction schemes for this artifact have been advocated at the volume resolution enhancement stage (as detailed in Section 6.4.2) or the large contrasting region removal step (described in Section 3.4.7) could have its mask dilated further.

Further confusion was observed at the borders of multiple sclerosis lesions (Fig 5.21), where the resolution enhancement process failed to ignore the diffuse borders and instead created surrounding rings of GM-like intensity. These regions could be excluded from processing or merely smoothed through the identification of lesions,

for example by examining connectivity of CSF regions with cortical sulci or by using a brain atlas as a CSF region template and flagging outliers (as advocated in Section 6.4.1). Prior knowledge of brain structure would also be extremely useful to remedy instances of extremely poor GM contrast (as shown in Fig 5.22), in addition to segmentation methods which are more robust to considerable overlap of components of the Gaussian mixing model such as employing the Bethe approximation to the Gibbs free energy and minimising the marginal posterior probability [Chen et al., 2005].

An area of processing demanding more formal investigation is the trade-off between sensitivity to small detail and ensuring high class content homogeneity in the neighbourhood clique scoring function (Eqn 5.18) and the ambiguity reduction schedule β_z; for example, this could prevent some of the spottiness in WM/GM partial volumes observed in Fig 5.21. Tuning these functions, or providing mechanisms for automatic training as performed by FAST and the unified segmentation method, would likely reap rewards in improved accuracy. When considering the application of deformable structural models (with proposals outlined in Section 6.4.1) to provide strong priors for region borders and high label homogeneity within those borders, then the neighbourhood clique scoring scoring function may need to be redesigned. The stage of processing at which the fractional content of voxels should be determined should also be further examined in future work. FAST's estimation in its post-processing – once voxels affected by PVE have been identified – may be more accurate than the label assignment in the E-step (Section 5.3.3) since it is purely intensity-based and can assign continuous class proportions. However, without neighbourhood context this type of estimation may be sensitive to noise.

Practical considerations such as memory limitations and processing time limit the types of label and parameter optimisation that may be considered, but the requirement for repeatable production of results stated in Section 5.3 excludes stochastic processing. Relaxation labelling [Rosenfeld et al., 1976] (employed in Section 5.3.4 as part of the E-step) was thus chosen for the task of label optimisation since it provides a wider spectrum of neighbourhood label interactions to affect the optimal configuration than the greedy ICM algorithm [Besag, 1986] at the penalty of increased processing time for calculating the extra clique potentials. To reduce the processing time, which is much slower than FAST but considerably faster than the implementation of the unified segmentation method provided, then the tolerance of its convergence criteria (stated in Section 5.3.6) can be made less strict at the penalty of less optimal solutions being reached. Alternatively, the ambiguity reduction schedule of the neighbourhood con-

text label prior – β_z in Eqn 5.14 – can also be tuned to achieve different performance. A quickly-decreasing β function leads to fewer R-step sub-iterations being performed and hence poorer label homogenisation, while starting with low β_0 results in less deviation from the minimum result of the unary r function (Eqn 5.16), and vice versa. In regions with low noise or high spatial frequency structures then quickly decreasing β_z and low β_0 should be preferred; these parameters could be tuned to optimise performance for specific ROIs. Improving the quality of the initialisation (described in Section 5.3.2) may also be a worthwhile strategy for improving accuracy at a low cost, for example by incorporating shaped and oriented neighbourhood weighting filters and thin structure priors in the objective function of FCM clustering (Eqn 5.6).

More efficiency concerns arise when processing high-cardinality neighbourhood cliques – 124 interactions for each label combination with a central element for a $5 \times 5 \times 5$-sized filter used at upsizing ratio $L = [2,2,2]$ – at the sub-voxel level in contrast to much lower-cardinality ones used in the literature (for example, just 2 in [Bricq et al., 2008] and a more common 26 for FAST). Therefore, if more complex optimisation processes are to be employed then they must be computationally efficient in the number of sub-voxels along with the neighbourhood size (both determined by L) as well as in the number of labels considered in M (determined by the number of fractional levels M^{frac}), as described in Section 5.4.10. Memory limitations in the R-step constrain the number of slices that can be processed at once on conventional desktop PCs, and in turn they limit the number of samples available for class parameter re-estimation (examined in Section 5.3.5) as well as the potential accuracy of this step. The efficiency of the M-step could be improved slightly by fixing the CSF class parameters to those obtained by sampling from the ventricles (located using a method such as that detailed in Section 3.4.7), but over the entire EM algorithm then search over 12 parameters – 2 means and 4 covariance elements each for both WM and GM – would still be required.

Message passing through belief propagation [Pearl, 1982] is an extremely interesting alternative to relaxation labelling since there is the potential to reduce memory requirements as well as to combine it with region-growing and contour-based structure boundary priors (discussed in Section 6.4.1). By initialising the list of sub-voxels to process in the middle of putative homogeneous regions and moving sub-voxels which have high spatial variation in their labelling (due to the presence of structure borders or noise) to the end of a priority queue, then by the time these sub-voxels are processed then more definite beliefs may have been established in their neighbours. The sequential processing may need to be streamlined by starting processing in multiple

geometrically-separated regions simultaneously and processing in parallel, as well as by setting flags for voxels that require no further action (such as in large uniform regions of WM or inside the ventricles) after a few iterations have occurred.

5.6 Summary

This chapter presents a new segmentation process inspired by the unified segmentation method [van Leemput et al., 2003]. As the culmination of the work in this thesis, following an investigation of data modelling assumptions in Section 5.2 then the methods described in Section 5.3 employed the resolution-enhanced volumes produced in Chapter 4 to simplify segmentation decisions that needed to be made as well as the thin structure maps and adaptive filters made in Chapter 3 to improve the quality of labelling around thin blood vessels and cortical sulci.

A previous analysis of FAST [Zhang et al., 2001] in Section 1.3.5 revealed weaknesses around sub-cortical and thin structures, which were also particularly evident for the unified segmentation method in Section 5.4.8. By using a soft – but still discrete – sub-voxel label set then a range of mixtures could be expressed at the voxel level while the appropriate labelling for partial volume structures could still be assigned at the sub-voxel level. In order to further improve the labelling quality then the constrained mixel dimension was enforced at the voxel level and relaxation labelling [Rosenfeld et al., 1976] provided a more deliberate labelling optimisation method than ICM [Besag, 1986].

The literature review in Section 5.1 revealed that testing on real and pathological datasets was less common than quantifying performance on a standard dataset such as the BrainWeb phantom, and the testing efforts made in Section 5.4 were more extensive and wide-ranging than those produced in previous chapters. Both quantitative and qualitative comparisons were made between the proposed methods and both FAST and the unified segmentation method, revealing improved labelling accuracy and robustness to noise except in a few specific cases, investigated in the Discussion (Section 5.5). Performance improved as components of the new segmentation process were added to a basic method, and the methods did not fail when applied to severely atrophied real brain data nor when exposed to multiple sclerosis pathology in the BrainWeb phantom (Section 5.4.9).

Some interesting topics covered in the Discussion (Section 5.5) included the possible use of the Bethe approximation to the Gibbs free energy and minimising the

marginal posterior probability [Chen et al., 2005] in order to improve WM/GM discrimination, as well as developing parameter training methods and applying further structural priors to inform the labelling where the intensity is not informative. The latter topics have been covered in more detail in the Conclusions (Chapter 6).

Chapter 6

Conclusions

6.1 Summary of findings and evaluation of the fulfilment of original objectives

The overriding aim of this thesis was to improve segmentation performance in thin structure and sub-cortical areas, which was achieved in most of the ROIs studied, compared to FAST [Zhang et al., 2001] and the unified segmentation method [van Leemput et al., 2003] in Tables 5.4 and 5.7, respectively. The most notable improvements in accuracy occurred when sub-cortical structure or many sulci appeared in an ROI, such as in the cerebellum (ROI 4, Fig 5.13). The segmentation of sulci was generally very successful due to their good intensity contrast, the high sensitivity and specificity of the thin structure maps produced, and the quality of the restoration of core sulcal intensities during resolution enhancement. However, the processing of blood vessels was hindered by their very poor resolution (especially in the through-slice direction for the LBC1936 volumes) compared to their thickness, and their relatively poor contrast with WM and GM in partial volume voxels. Sulci have a plate-like structure and usually have a greater thickness; hence they were not so badly affected and could afford to be thresholded at lower values, increasing sensitivity without significantly damaging specificity. Consequently, additional measures may have to be taken such as explicit vessel structure modelling or even co-registration of MRA images for more accurate segmentation of blood vessels to occur.

The objectives of the thesis were derived from the desire to improve segmentation performance while also forming a simpler segmentation problem. Several modelling conditions must be fulfilled in order to form the simplest one possible:

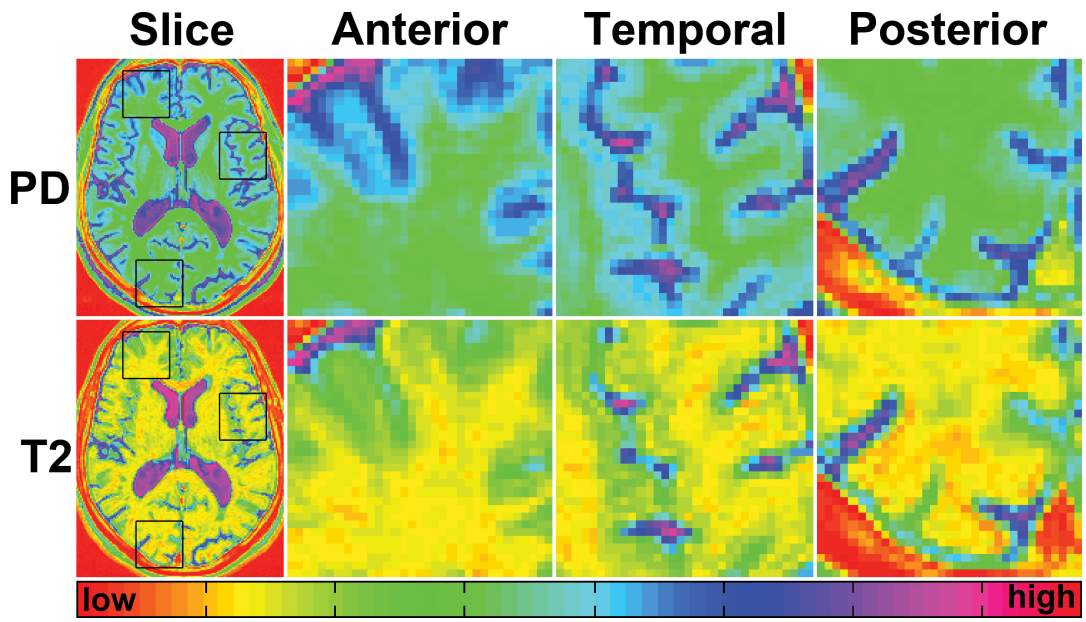

Figure 6.1: **Spatial variation in GM intensity**. **First column**: T_2 and PD slices on which ROI 5 is located. **Second column**: Zoomed-in anterior region. **Third column**: Zoomed-in temporal region. **Fourth column**: Zoomed-in posterior region.

1. The number of different tissue types should be as small as possible to minimise confusion between them (or their mixtures), and a simple model of intensity formation through class mixing within voxels should hold;

2. The features on which segmentation is based should be distinct for each class, the pure classes should be well characterised in datasets by a particular distribution, and each tissue class should possess high intra-class homogeneity (low covariance for Gaussian distributions);

3. There should be negligible bias field effects and noise (intensity non-uniformity) as well as little pathology or artifacts;

4. There should be strong and effective priors for anatomical structure boundaries or for the localisation of class content.

Considering the second and third issues, each chapter's results have been affected by bias field effects (first described in Section 1.3.3. GM in anterior regions was better contrasted from WM and CSF than in posterior ones (as evident by the spatially-variant contrast in Fig 6.1), leading to substantial confusion between WM and GM in ROIs 2 and 10 (observed in Section 5.4.6). It could be argued that improvements in segmentation accuracy using the proposed methods were caused by assigning bias-affected

regions fractional content labels, preventing the class parameter estimation from being too badly affected, but the effect was largely the same for FAST (as shown in Fig 5.11). Running FAST with its bias correction steps enabled did not produce significantly more accurate results (listed in Table 7.3), with ROIs 2 and 10 remaining the only ones improving upon the accuracy of the results of the proposed methods presented in Table 5.4. Nevertheless, such spatial variation in intensity does demand correction to improve accuracy in poorly-contrasted regions (such as the posterior areas) and an extension to existing bias correction methods has been suggested in Section 6.5.3. Incorporation of a bias correction step would also allow more direct comparisons with the full processing of FAST [Zhang et al., 2001], for which characterisation of the bias field is a substantial component of its segmentation pipeline.

All the methods described were able to process dual-echo T_2-PD MRI structural brain volumes (the first main objective listed in Section 1.4) and to exploit the tissue contrasts. Each of the methods chapters then fulfilled their main purpose (the second to fourth objectives). The thin structure tissue content priors were trained to produce a very high positive predictive value and hence were valuable priors for later segmentation processes since traditional atlas-based priors are not sufficiently relevant for these thin structures, which have unique configurations between individuals. They also helped to simplify the segmentation process through the thin structure restoration performed in the resolution enhancement phase, forming more pure tissue regions. The adaptive filters produced proved useful later for preserving thin structure extent and contrast in both the resolution enhancement and segmentation stages.

Resolution enhancement aimed to enhance the quality of the data in order to make it more amenable to accurate segmentation, through adaptive smoothing and PVE reduction. This process was successful at restoring the pure tissue intensities of the thin structure mapped in the previous chapter, and showed good congruence of edges when comparing resolution-enhanced results to higher resolution volumes (Figs 4.12 and 4.13). However, where contrast was poor – particularly between WM and GM – or where thin structures were not mapped, then little enhancement or even blurring occurred and the localisation of tissue borders was still sensitive to noise. This issue may be tackled by structural modelling and by applying deformable atlas-based priors (examined in Section 6.4.1), or alternatively by noise removal in pre-processing (investigated in Section 6.4.2), to provide accurate localisation of these boundaries while still providing robustness to noise and not requiring more data.

The main objective for the segmentation process was to allow rapid variation across

region borders and thin structures but to prefer homogeneity in pure tissue regions. This was achieved by the combination of many components in the careful design of an energy function (Eqn 5.15) that incorporated an intensity-based component, the weighting of neighbouring label data using the adaptive filters, and the usage of tissue priors. A very important component was the clique potential scoring function that preferred exact label matches but provided a continuous response to partially-matching labels; in addition, a penalty the enforced the constrained mixel dimension at the sub-voxel level. In combination with the output of the resolution enhancement process – where noise had been smoothed and PVE had been reduced – then rapid variation in the labelling was uncommon at the sub-voxel level and so the design of the clique interaction scoring function and the relaxation schedule β_z could fulfil the objective at the voxel level. Training methods (as examined in Section 6.3) or automatic re-evaluation of β_z, as performed in [van Leemput et al., 2003], to precisely adapt the relaxation schedule to the data at hand would likely provide valuable improvements to the proposed methods' accuracy, as would the usage of segmentation methods which are more robust to overlap of the WM and GM intensity distributions (such as employing the Bethe approximation to the Gibbs free energy and minimising the marginal posterior probability [Chen et al., 2005]).

Perhaps the most important result arising from the demonstrations of the methods proposed in this thesis is the gradual improvement of results presented in Section 5.4 (with a build-up pictured in Fig 5.9) as modifications were added to a basic segmentation method. This finding suggests that volume resolution enhancement, thin structure-specific processing and more deliberate labelling optimisation could be added – even if only as optional features – to existing segmentation tools with beneficial effects on their accuracy.

Throughout the analyses of the demonstrations' results in Chapters 3 and 5 then regions of the *other* class assigned by Dr Farrall to a small number of tissue types not fitting the four main tissue classes have been ignored. Their similarities in intensity to the main tissue classes necessitate the application of anatomical structure priors, as advocated by works registering probabilistic anatomical atlases to classify these regions i.e. [Ashburner and Friston, 2005]. Later in this chapter (Section 6.4.1) plans for adaptive structural models could allow the correct identification of glial matter, membranous tissue like the septum pellucidum, and the choroid plexus; this step will likely lead to further increases in labelling accuracy and also improve the methods' clinical utility.

All the methods described in this thesis do not employ stochastic processing, so the results can be exactly reproduced when the processing is repeated on the same data using the same parameters: this is an important requirement for good precision, one of the key evaluation metrics identified in Section 1.3.5 and the fifth main objective. Furthermore, formal training procedures were specified for a number of important parameters (including the range of thin structure scales in Section 3.4.6) so that the likelihood of picking the same parameters between different observers is increased. Another factor aiding this process is that the purpose of the parameters has been documented well in the methods chapters and their summaries, meaning that their selection is less likely to be confused.

The wide range of testing performed (fulfilling the final main objective) showed the robustness of the proposed methods. In no example studied was a degenerate segmentation result produced, where two classes were combined or one class existed mostly as partial volume in pure tissue regions. The range of data studied included:

- Aged brain volumes, with thick slices, in the LBC1936 cohort. This dataset had a ground truth labelling provided by an expert. These volumes were also downsampled for evaluating the entire segmentation method;
- Severely atrophied brain tissue from one LBC1936 subject, which was only used in Chapter 5 for evaluation of the entire segmentation process;
- BrainWeb data with no bias but a range of noise levels, and isotropic voxel dimensions. This phantom data naturally had a ground truth supplied;
- BrainWeb data with multiple sclerosis lesions. This dataset again had isotropic voxel dimensions and a supplied ground truth, and was only used in Chapter 5 for evaluation of the entire segmentation process;
- A young brain volume acquired with thick slices, and used during evaluation of the thin structure detection in Chapter 3 as well as the build-up of the segmentation process in Section 5.4.5;
- A young brain volume acquired at two different resolutions with thick slices, used for evaluating the resolution enhancement process in Chapter 4.

The usage of the real LBC1936 dataset was particularly worthwhile since it meant that the methods were not evaluated solely on the BrainWeb phantom, an unfortunate

quality of many segmentation papers in the literature. In clinical practice MRI is more likely to be performed with a high slice thickness (to improve SNR and reduce acquisition time) on older patients and so this dataset is representative of that population and typical acquisition settings. Furthermore, the inclusion of patient with severe brain atrophy in this dataset demonstrated the robustness to extremely poor contrast and some pathological tissue. In future developments, the LBC1936 dataset will likely be of further use due to its inclusion of T_1-weighted and DTI volumes in the examination. These could provide other useful tissue contrasts and structural information if very accurate co-registration can be employed.

Considering some of the less important objectives also stated in Section 1.4, the viability of using Gaussian distributions to model pure tissue intensity spread (as commonly assumed in the literature) as well as the mixel model were examined in Section 5.2 on the LBC1936 data. The results from the normality tests did not point to the data being Gaussian-distributed, but as noted in Section 5.2.2 the small sample size and lack of removal of the bias field may have been factors in the rejection of Gaussianity. In evaluating the efficiency of the described methods, the running times for full volumes recorded in Sections 3.5.5, 4.3.4 and 5.4.10 were lengthy but did achieve the goal of being expressed in the order of hours rather than days. More importantly, the computational complexities were linear in the data size, meaning that scaling up to higher resolutions should not be a substantial problem, and the code was able to run on a single desktop computer.

While the total of processing times were slow compared to the running time of FAST – often taking less than half an hour per volume – it was much faster than for the unified segmentation method, which took around ten minutes to process a single slice due to the MCMC sampler involved in optimising the MRF. However, the implementations were not very efficient and suggestions for improvements have been made in each of the preceding three chapters. Furthermore, processing time is generally a much cheaper and less troublesome resource than scanning time and it scales with the ever-improving hardware. In terms of clinical practicality running time is a more serious issue, but it can be alleviated with divide-and-conquer processing on cluster and grid computing platforms which are becoming more widely available, especially to research establishments.

Revisiting the idea of simplifying the segmentation problem, there exist interdependencies between the proposed system's components of thin structure detection, deformable atlas-based priors, volume resolution enhancement and bias correction. Once

structure boundaries are identified then internal intensity variation can guide characterisation of the bias field and volume resolution enhancement becomes a reconstruction process, dependent on the class content and class mean products of segmentation (similar to [van Leemput et al., 2003]). These boundary contours might be positioned based on metrics derived from labelling (as used in FreeSurfer [Dale et al., 1999][Fischl et al., 1999], and investigated further in Section 6.5.4) and thin structure detection; segmentation itself will also need to work with bias-corrected resolution-enhanced data. The complex dependency loops that form consequently prompt the design of a unified (rather than sequentially-applied) framework for the segmentation system's components.

The most well known of these types of frameworks (i.e. [Pohl et al., 2005][Fischl et al., 2004]) was proposed in [Ashburner and Friston, 2005] to combine bias correction, spatial normalisation to a tissue content prior atlas and hard segmentation in a single model[1], making it robust to large errors common in processing noisy, low contrast and low resolution fMRI data. The authors argued that forming a unified energy minimisation problem with contributions from each of these components and using ICM [Besag, 1974] to optimise their parameters in a round-robin style would allow more complete – though likely slower – optimisation because of the interdependencies among the parameters of different components. The framework proposed in the previous paragraph contains two additional components as well as more complex atlas registration, and so the simplistic optimisation style would need to be preserved in order to be computationally viable. Despite the large amounts of development time required for the framework to reach practical fruition, it appears the most satisfying method of correcting for segmentation errors where the tissue content cannot be accurately determined from the intensity presented. As part of a unified segmentation process, reconstruction of the intensities following segmentation – as seen for the putamen in Fig 6.2 (marked with a red arrow) – violates the intensity conservation property of diffusion and hence resolution enhancement will require either gradual repair of the original data to match current internal anatomical models, or diffusion-based and labelling-reconstruction-based processing modes that could be switched between as the bias, anatomical structure and tissue class feature models improve. More detailed plans for such a unified segmentation method, however, will not be covered in this thesis and remain an intriguing prospect for future research.

In the remainder of the this chapter, the first two sections will examine practical

[1]This model has been implemented in the Statistical Parametric Mapping package.

issues encountered with testing, parameter training and rater evaluation. Subsequently the groundwork for the application of deformable models to brain segmentation as well as the correction of noise and "ghost" artifacts will be presented (Section 6.4). Next, further applications of the proposed methods are discussed in the fields of DTI, perfusion MRI, bias correction and multi-modal (or multiple image weighting factor) co-registration in Section 6.5. Finally the existing and planned publications of work described in this thesis will be referenced (Section 6.6).

6.2 Commentary on datasets and the range of testing

Chapters 3-5 examined changes in performance of the proposed methods, either through varying their parameters or by sequentially applying processing modifications to a basic method to induce an increase in performance and thus prove their viability. Only in Chapter 5 was performance directly contrasted against different, rival methods; if time allowed, a more complete testing strategy would have at least quantified the differences of the thin structure detection method with the flux-maximizing-flow-based algorithm [Descoteaux et al., 2008], resolution enhancement with another single-frame image super-resolution method (i.e. [Jiji and Chaudhuri, 2006]), and the entire system with other soft segmentation tools (such as [Bricq et al., 2008]). Re-compilation of the source code of FAST [Zhang et al., 2001] to incorporate implementations of the proposed resolution enhancement (since using the resolution-enhanced volumes as input did not produce satisfactory results) and thin structure detection stages into its processing pipeline, with subsequent testing against the original tool, would give more weight to their billing as beneficial modifications.

With the exception of the BrainWeb phantom, each dataset processed was representative of MRI data likely to be acquired in clinical practice from young and aged patients. This caused method development and testing to be particularly challenging due to the high slice thickness which hindered the detection of some region boundaries and reduced the intensity extrema of many thin structures (as noted many times in Chapter 3) through PVE; furthermore, poor contrast between WM and GM – particularly in occipital regions for both image weighting factors – gave the potential for degenerate segmentation results and complicated class parameter initialisation. The amount of expert-labelled data available was fairly small but labelling each of the 10 LBC1936 dataset ROIs took a considerable time, and in retrospect some ROIs with duplicate anatomical content could have been replaced with larger ROIs requiring just

binary labelling of thin CSF or BG structures (for testing work in Chapter 3) which should be fairly rapid to produce.

In the literature, standard datasets – with a ground truth supplied – enable comparisons to be drawn between different methods (as summarised for BrainWeb in Table 5.1). While the Internet Brain Segmentation Repository[2] datasets may not be entirely suitable for testing a soft segmentation method due to their hard labelling, they are sourced from real MRI acquisitions containing more realistic models of anatomy than BrainWeb and hence they have often been selected as the standard dataset for quantitative round-ups; for example, many of the papers listed in Table 5.1 list accuracy statistics for these volumes. Recognising the need for a MRI brain volume phantom derived from real MRI acquisitions, in [Aubert-Broche et al., 2006b] a 20-volume dataset was produced with tissue content maps for many tissue classes in the head including WM, GM, CSF and blood. At present only T_1-weighted volumes can be generated, but if volumes with other image weighting factors become available in the future, then further testing on work in this thesis could substitute this dataset for BrainWeb.

Assessing the transferability of the methods across different acquisition protocols, scanners, subjects and aspects of pathology would also be desired – though the amount of expert-labelled data required for quantitative evaluation would be restrictive. On-line segmentation data corpora (such as the facility recently described in [Shattuck et al., 2009]) go some way to tackling the issue of organising multi-centre co-operation, but testing robustness to many types of pathology[3] other than multiple sclerosis lesions and atrophy that could be encountered requires access to a range of patient data. A more realistic proposal would be to study age-related degeneration apparent in the LBC1936 dataset which includes reductions in GM thickness, widening of CSF-containing bodies, and the development of small blob-like oedemic features in the GM (as observed in Fig 5.10).

Structural changes and intensity differences induced by pathological tissue will challenge the proposed segmentation methods and their basic modelling assumptions (examined in Section 5.2), though some will be easier to account for than others. For example, the multiple sclerosis lesions examined in Fig 5.21 consist mainly of small blobs of oedema in the WM, which could be simple to identify by adjusting the thin structure detection method to respond to hyperintense blobs (high S_{new}, low R^B_{new},

[2] Available from the Center for Morphometric Analysis, Massachusetts General Hospital, USA at `http://www.cma.mgh.harvard.edu/ibsr/`.

[3] Visualisation of many types of brain disease pathology can be found at `http://radiology.uchc.edu/eAtlas/nav/msBrain.htm`.

$contrast$ = bright). Like multiple sclerosis lesions, tumours will also be problematic for resolution enhancement due to their non-uniformly diffuse borders, and since they contain many poorly-formed small blood vessels and oedema then the resulting PVE and rapid variation in tissue content will complicate labelling. Consequently, prior identification of specific pathology (such as tumours) with removal of these regions from the parenchymal processing mask may be the most robust option. More general pathology detection, assessing the deviation from a normal brain template, may complicate the simple modelling assumptions valid in normal brain tissue.

An aspect of testing that was severely constrained by hardware constraints was the the upscaling ratio L for the volumetric data which was always set to $[2,2,2]$. At higher ratios then matrices are no longer processed solely in memory but instead need to be paged from the hard disk, vastly increasing processing time; the same reasoning explains why only a few slices surrounding each ROI were processed, as opposed to the entire volume. Some testing has been performed at other upsizing ratios including $L = [3,3,3]$ with the other parameter values remaining the same as specified in Section 5.4, and the segmentation results are presented in Figure 6.2. While the resolution-enhanced volumes appeared more visually attractive with sharper boundaries there was no substantial increase in labelling accuracy, suggesting that some parameter tuning may be required.

One criticism of the acquisition protocol is that the TE could have been set to provide a greater distinction between T_2 and PD tissue contrasts (especially between WM and GM). Multi-channel processing increases robustness to noise but almost as valuable is the combination of these contrasts to ensure good tissue class separation in the intensity space, and the WM/GM contrast (normally good) in the PD-weighted data did not make up for its poor contrast in T_2. However, the LBC1936 data was originally acquired for another study and so the segmentation concerns relevant to this thesis may not have been paramount in the scan sequence's design. It can be observed that the young brain and BrainWeb datasets possess much improved contrast for these classes, especially in occipital regions (red arrows in Fig 2.1) – though they represent ideal acquisition conditions not usually realised in clinical practice and a phantom, respectively. Co-registering T_1 data and including this image weighting factor as an extra dimension in the feature space (as performed in [Anbeek et al., 2005]) to exploit its excellent WM/GM contrast may help to improve these classes' distinction, but unless perfect alignment can be guaranteed then its usefulness may be relegated to a spatial prior for WM/GM boundary locations.

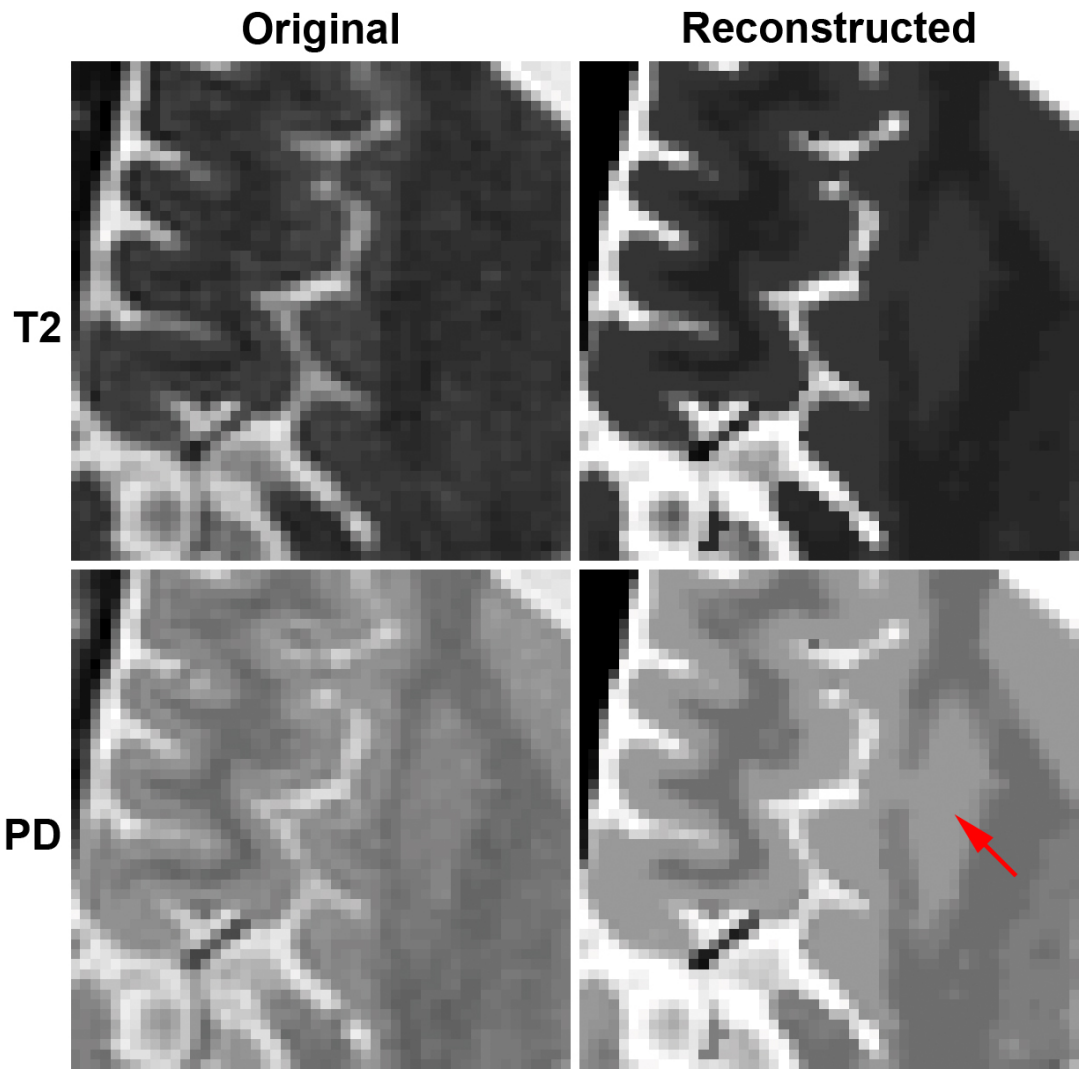

Figure 6.2: **Reconstructed intensities after using the full segmentation process with a higher upscaling ratio**. **Left column**: T_2 and PD data from ROI 5. **Right column**: reconstruction of the T_2 and PD values after the original data were processed using all the methods described in Chapters 3-5 with a $3 \times 3 \times 3$ upscaling ratio, and then the labelling was downsampled to the original resolution (Dfn 5.5). The values were then formed from the linear combination of the estimated class content proportions and the class means output from segmentation (Dfn 5.2.1). The coloured arrow is referenced in this chapter's introduction.

6.3 Proposals for training frameworks

Some of the parameters required for the proposed methods in Chapters 3-5 are relatively easy to set (such as the desired number of tissue classes), a few are limited by hardware or time constraints (like the upscaling ratio or diffusion timestep), while for some others training procedures have been outlined (i.e. Proc 3.4.1 for estimating the range of scales relevant to thin structures). The objectives for optimising the remaining parameters are outlined below, and their practical implementation may involve replacement of qualitative goals with quantitative targets. Training should always be performed with the assistance of a radiologist to ensure that gross errors (with respect to the anatomy) do not occur.

- The primary performance factor of the thin structure detection stage is the high PPV of the thresholded maps B (Dfn 3.4.5) – false positives should be minimised, and so underestimation of non-zero fractions of the ground truth class content will be likely. The true positive rate can then be maximised as a secondary objective. When investigating the correctness of the principal normal directions, they should vary smoothly along blood vessels and across the thin dimensions of sulci, but should also respond to tight changes in their orientation.

- The RD phase of volume resolution enhancement must principally reduce PVE at tissue borders, causing an increase in pure tissue class membership (as observed in Fig 4.15) as well as a qualitative sharpening effect at structure boundaries which is especially visible around thin structures, WM/GM interfaces and at the edges of the ventricles. Smoothing of homogeneous tissue regions in the RAD phase should not lead to blurring of these established boundaries, and in both phases pure tissue intensities should be restored at the core of thin structures identified in B. It is important to monitor that the sizes of structures are not dilated or contracted beyond sub-voxel changes expected through PVE reduction, by examining the error between the downsampled result and the original data.

- FCM clustering must correctly sample the BG class to produce summary statistics similar to those of intensities found at the core of the largest veins, and the membership maps for all four tissue classes must be fairly realistic. An easily-checked criterion would be to visually assess whether the cores of the cortex, ventricles and large WM tracts possess very high memberships of GM, CSF and WM, respectively.

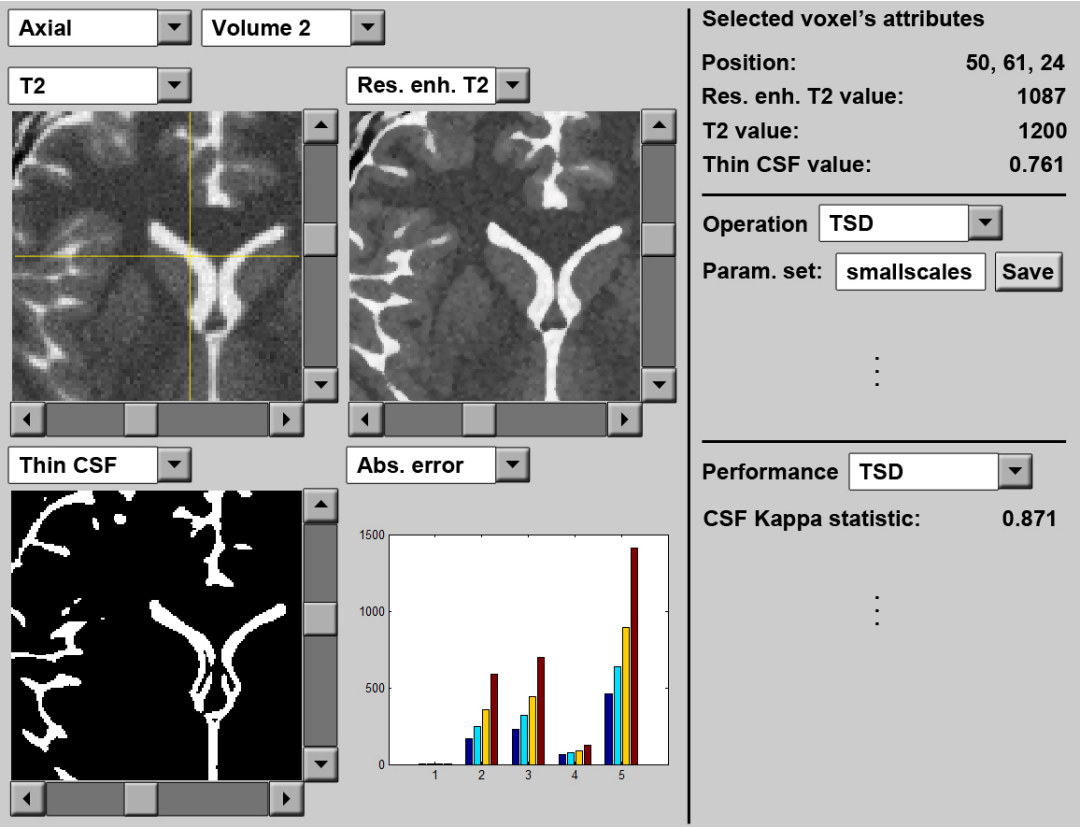

Figure 6.3: **Mockup of a parameter training interface**.

- The segmentation stage may have different objectives depending on the source data and the particular application of the result but will usually demand either the maximisation of a labelling accuracy measure (such as the absolute error or κ-statistic score) with respect to a supplied ground truth, the correct delineation of a particular structure (assessing an overlap measure such as the Dice metric [Zou et al., 2004]), or the generation of realistic tissue class parameters (compared with the parameters derived from manual sampling of pure tissue regions and the application of the M-step in Section 5.3.5). Qualitative optimisation will involve examining more subjective concerns such as robustness to noise and how well thin regions or tissue boundaries are preserved in the labelling.

Since the relationships between components of the proposed methods (illustrated in Fig 1.23) are not trivial, a semi-automatic software tool for parameter training using these guidelines would be beneficial. Its main purpose would be to allow different parameter sets to be batch processed and their results to be compared and contrasted across a range of metrics in a time-efficient and precise way. A mock-up of a parameter training tool interface is pictured in Fig 6.3 showing feedback occurring at several

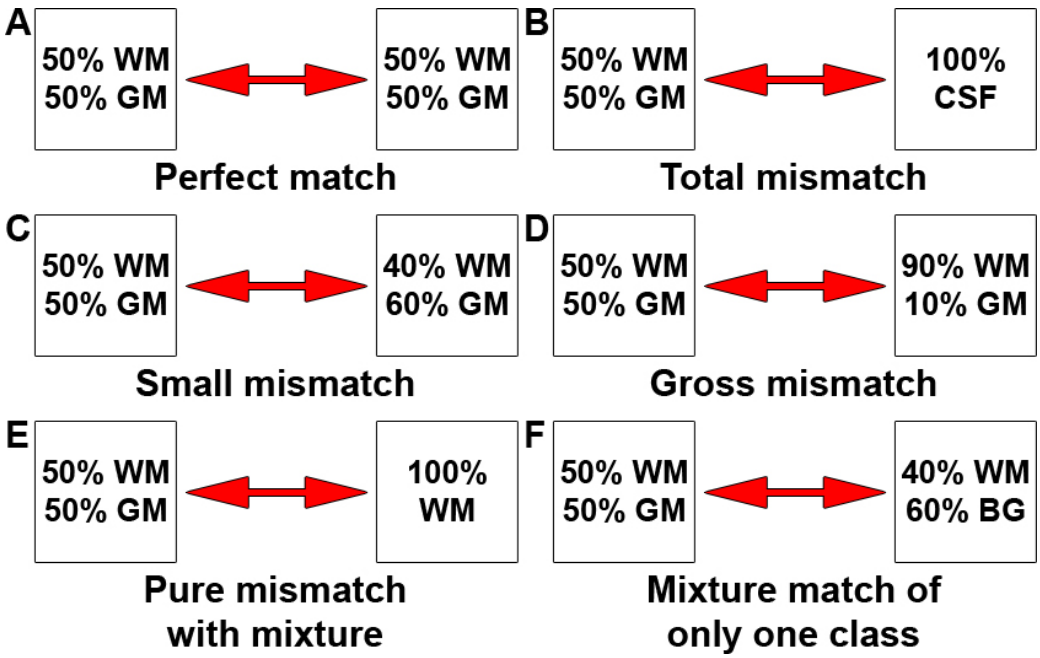

Figure 6.4: **Illustrated cases of soft label matching**.

levels including single voxels (lists of values), pictured slices or ROIs (using an array of visualisation options), and volumes (large-scale statistics). Ideally expert-labelled data will be available to provide a ground truth, as advocated in [Nyul and Udupa, 2002], to aid some subjective decision-making with quantitative metrics during parameter training.

The surrogate of truth produced for the LBC1936 dataset in this thesis was produced in a single pass performed by a single expert, and so this result will be affected by rater bias, human error and current working conditions. In order to improve this surrogate of truth then repeated labellings can be consolidated and results can be unified from several experts, requiring a labelling consensus formation system. The Simultaneous Truth And Performance Level Estimation (STAPLE) [Warfield et al., 2004] framework has been designed for this purpose; to extend STAPLE to work with soft segmentation results then the hard label matching (effectively implemented using a delta function) must be replaced with a case-based or continuous scoring function to penalise less serious label overlap errors (such as those pictured in Fig 6.4C and E) less heavily than more serious ones (such as Fig 6.4D and F) where the class constituents are completely different. A further application of this type of case-based label matching system could be in the formation of isosets of voxels for the analysis and visualisation of segmentation errors.

Related to the tasks of parameter training and error analysis for automated tools is the education of human raters to perform more similarly to selected experts. In both cases then the characterisation of trends in label confusion may best supplement the metrics used in Section 5.4. The diagonal elements of the conditional probability matrix of a rater's labelling given the consensus point to the probability of perfect matches, but clustering the more interesting off-diagonal elements (mismatches of varying degrees of seriousness) is complicated by the ordering of the labels referenced by the rows and columns of this matrix. To provide a clear relationship between adjacent matrix elements and allow clustering to take place, then either the matrix must be remapped to a $2C$-dimensional space (where C is the number of classes) or an adjacency matrix can be employed, as used in spectral clustering [Shi and Malik, 2000]. If the soft label set is particularly fine-grained then it may have to be coarsened so that a large number of small and uninformative clusters are not prescribed.

6.4 Improvements to the proposed methods

In this section two improvements have been outlined which may improve the accuracy of the results presented in Chapters 2-5, but were not implemented due to lack of development time. The most important is the incorporation of contour-based adaptive anatomical models (Section 6.4.1) which will alter definitions of structural continuity and homogeneity as well as provide useful indicators of the properties of neighbouring regions. Secondly, Section 6.4.2 investigates corrective measures applicable during the resolution enhancement stage for strong patches of noise as well as for dark artifacts existing close to large regions of CSF.

6.4.1 Incorporation of atlas priors and model-based constraints

Tissue content priors supplied by prototype or probabilistic atlases (such as [Evans et al., 1994]) have been rejected for use in this thesis as they cannot accurately represent individual brains with unique and complex small-scale structural variability [Roland and Zilles, 1994]. In this section the potential benefits of using deformable atlases (i.e. [Gee and Bajcsy, 1998]), which can be elastically transformed to reflect individual anatomy, for modelling brain structures will be investigated. In particular, blood vessels and cortical sulci relevant to thin structure detection (Chapter 3) demand notions of thickness, continuity and branching for the generation of realistic networks.

The pitfalls of most deformable models are largely the same: user intervention may be required to ”polish” results or initialise processing, definition of structure shapes and their modes of variation may be difficult and time-consuming, there may be poor representation of short or branching structures, and processing time will increase to incorporate their non-rigid registration (described in Section 1.3.2 and further examined in Section 6.5.4) by assessing the matching of features including intensity, intensity gradients, and the thin structure measure T_{new} (Dfn 3.4.4). Despite these drawbacks, contour-based representations can model PVE as structure boundaries cutting voxel volumes – providing valuable tissue mixing or purity priors – and constraints on the topology of the region boundaries may offer robustness to noise. The latter factor is typically problematic for local neighbourhood label interaction models such as MRFs when faced with high frequency noise at edges. These representations can also inform many aspects of structure homogeneity over different spatial scales:

- The local uniformity of cortical thickness;
- The relatively consistent shape and existence of sub-cortical structures;
- The tubular shape and continuity of the blood vessel network;
- The plate-like shape of sulci and continuity with other sulci, the meninges or the ventricles;
- Dependencies in shape and tissue types between neighbouring regions;
- The heterogeneity of tissue content in some sub-cortical structures (such as striation of GM with WM).

Building full-brain models of anatomical structure requires several different components (pictured in Figs 6.5 and 6.6) and since the structure models will be contour-based then they will not be sensitive to the interpolation methods used and hence are more viable targets for co-registration than by using other types of brain atlases or incorporating additional image weighting factors. Firstly, active shape and appearance models [Babalola et al., 2008] can capture representations of regional texture (in the case of structural MRI, the intensity from each image weighting factor) as well as the modes of shape variation of border control points, and they have already been applied to model sub-cortical structures including the ventricles and putamen (Fig 6.5B). Their main benefit will likely be improving labelling in poor contrast conditions, which is often the case for sub-cortical GM in T_2-PD volumes. Sub-cortical structures classified

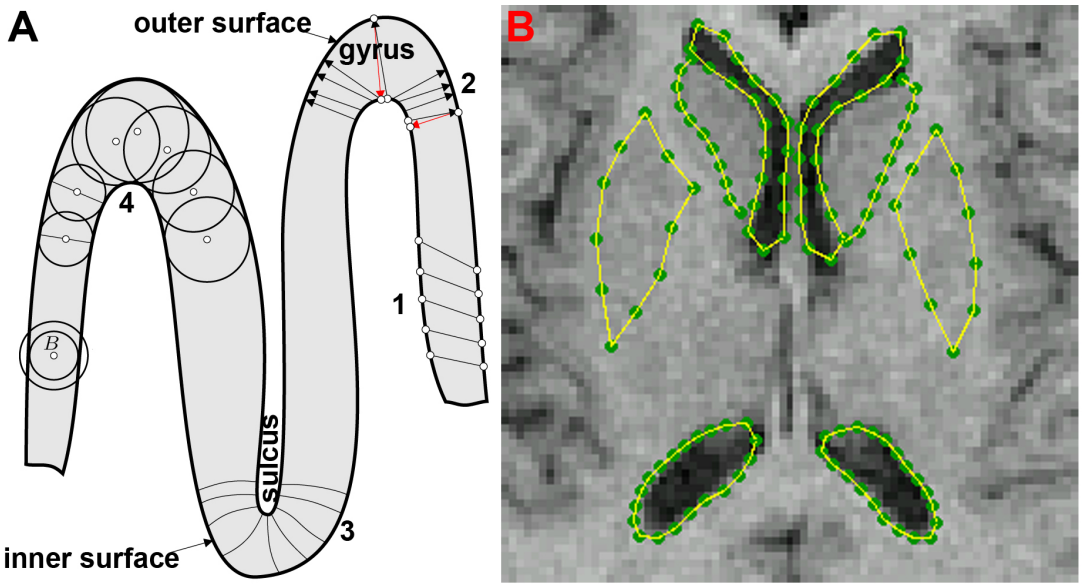

Figure 6.5: **Contour-based modelling of cortex and sub-cortical structures**. **A**) Coupled spline model of the inner and outer surfaces of the cortex, with several methods for assessing thickness constraints: corresponding points on coupled surfaces (1), cloest points (2), orthogonal flow lines (3), and inscribed spheres (4). Illustration modified from [Thorstensen et al., 2008]. **B**) Active shape model of a sub-cortical region incorporating the putamen, caudate and ventricles. Image reproduced from `http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/Models/pdms.html`.

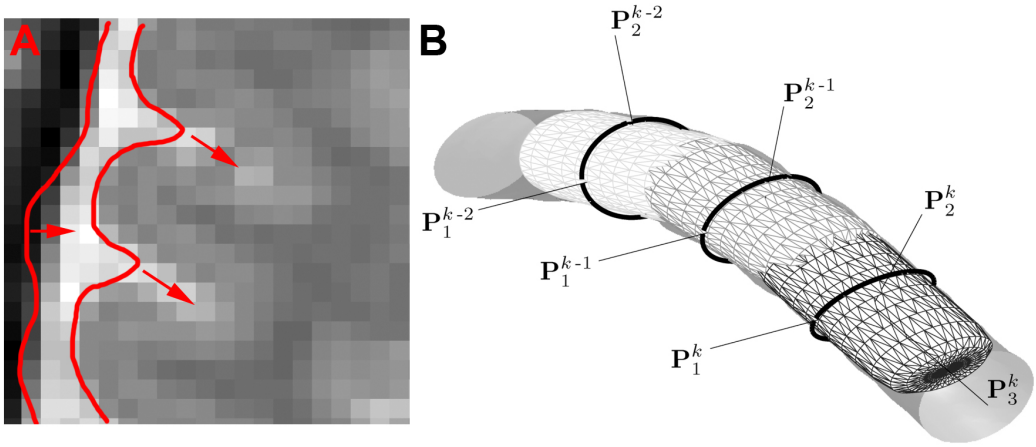

Figure 6.6: **Conceptual illustrations of additional structural constraints for thin structure detection**. **A**) Front propogation model for CSF structures. **B**) Fitting of superellipsoids (as an example of a 3D shape primitive) after medial axis generation. k is a position index for points **P**, and the subscript number denotes the ranked principal axis index of the accompanying line. Illustration reproduced from [Tyrell et al., 2007].

as *other* by Dr Farrall (noted in Section 5.2.1) could be identified using these models despite their lack of distinctiveness in their intensity from other tissue classes, since they still possess local intensity contrast from other neighbouring tissue regions.

The variability of cortical folding prevents a set number of border control points being appropriate for every brain's cortex, prompting the application of a coupled-spline model of cortical thickness (i.e. [Bosc et al., 2003] pictured in Fig 6.5A) instead, which can denote the inner and outer surfaces and account for thinner GM in the occipital lobe. One option for initialising cortical modelling is to surround the thin CSF structure map B^{CSF} (Dfn 3.4.5), though sulcal interruptions in highly-convoluted cortical regions must be carefully handled. For anatomy within the cerebrum, WM may be easier to model as it could be considered "filler" between GM structures within the brain parenchyma, and the skull and meninges can be appreciated as layers surrounding the cortex.

Cortical sulci and blood vessels are much smaller in scale, meaning their resolution poor in clinical MRI, and so ensuring continuity of the representation in the face of noise or sulcal interruptions is essential. Valid approaches (pictured in (Fig 6.6) include some front propagation methods [Ruan et al., 2000][Shen et al., 2005], active contour models (or snakes) [Lorigo et al., 2001][Goualher et al., 1999], and methods relying on the generation of medial axes[4]. The latter methods usually flag peaks in

[4]Medial axes are also known as centrelines or skeletons in the literature.

the principal direction of maximal local curvature as ridges [Smith, 2002] or creases [Koenderink and van Doorn, 1993] and then employ a shape tracking model along the extracted medial axes [Fritsch et al., 1995][Aylward and Bullitt, 2002]. This approach is preferred since sites discrete medial atoms sited along these axes can hold the proposed dimensions (perhaps derived using the local eigenvalues λ of the Hessian and the optimal scale in B, computed in Chapter 3), orientation (related to the principal normal direction in the local eigenvectors υ of the Hessian) and rate of widening or narrowing of envelope primitives such as maximal inscribed spheres [Antiga and Steinman, 2004][Hua and Yezzi, 2007] and super-ellipsoids [Tyrell et al., 2007]. There are other applications of these properties in the characterisation of the structures' shape statistics and also in shaped and oriented filtering for thin structures (Section 3.4.4).

Such cores-based methods have already been the target for establishing formal structure crossing, branch identification and termination criteria [Fridman et al., 2003] to ensure complete and well-formed results. To aid the generation of the medial axes then the Voronoi tessellation of thresholded thin structure sub-voxels in B or the minimum-cost path [Wink et al., 2004] through the thin structure measure T (Dfn 3.4.4) could be used.

6.4.2 Correction of noisy voxels and ghosting

Voxels which are affected strongly by noise may cause errors at each stage of processing: they can be identified as contrasting regions if they appear in clumps with the appropriate profile during thin structure detection (Chapter 3), they can cause fluctuations in in the local Gaussian-filtered gradient and be preserved during resolution enhancement (Chapter 4), and the conditional probability of the intensity may be strong enough to cause high-frequency variation in the labelling during segmentation (Chapter 5). Most of this section concentrates on the correction of noisy voxels' intensities at the resolution enhancement stage; a special case of correcting a long, dark artifact appearing parallel to the edges of large hyperintense CSF regions (named "ghosting" in Section 3.5.1) is tackled at the end of this section.

One strategy for correcting noisy voxels involves their detection as local outliers and modifying resolution enhancement to prefer smoothing in these regions: by adaptively increasing the noise threshold k used in the diffusion co-efficient (Eqn 4.17), the forward-time smoothing part of the co-efficient can be made more dominant. The voxels flagged in the thin structure map B (Dfn 3.4.5) should be removed from this out-

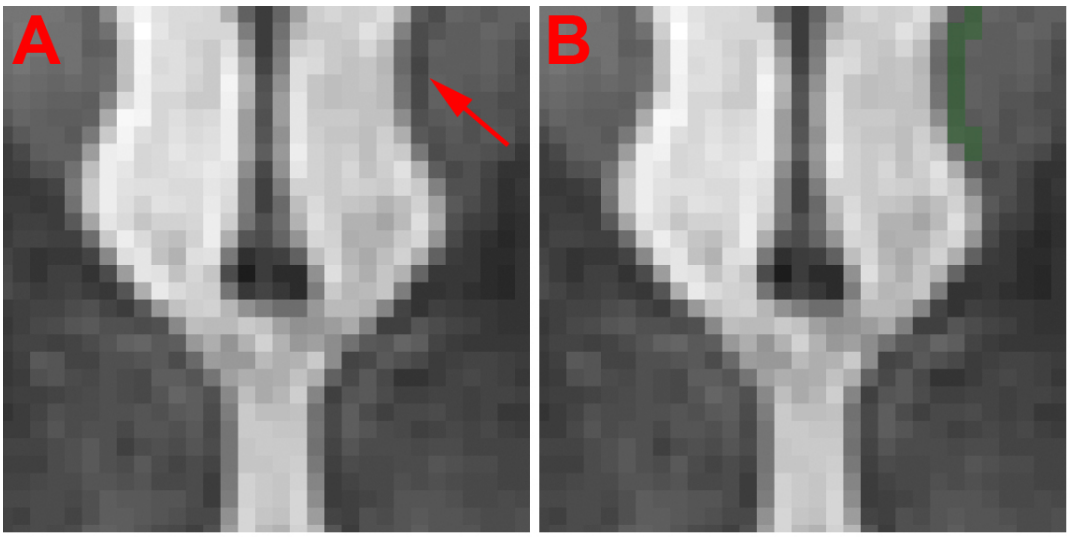

Figure 6.7: **Simple method for identification of ghosting**. **A**) Axial slice from a T_2 volume with a ghost highlighted by a red arrow. **B**) Identification of the ghost by its path parallel and close to CSF region edges, as well as by its lack of connectivity to the rest of the blood vessel network.

lier set to prevent them from being processed as such. Some simple outlier detectors include filtering based on the rank ordering of local intensities (similar to that used in [Salvado et al., 2006]) and the investigation of the highest frequency wavelet coefficients following a wavelet transform, though these approaches may lack specificity.

A more intriguing and preferred method for the correction of noise is non-local means [Buades et al., 2005][Buades et al., 2006] pre-processing, which is based upon global data statistics. Since images (or equivalently, large brain volumes) usually offer redundancy and self-similarity in patches – implemented as Gaussian windows – the weighted content of these patches can be averaged to remove small local fluctuations. An implementation for MRI volumes in [Manjon et al., 2008] produced impressive smoothing results but the degree of preservation of thin structures was unclear, and so regions flagged in B should again be processed very carefully or even excluded from processing. Furthermore, the effect of non-local means on PVE at the boundaries of tissue regions as well as the sensitivity to the neighbourhood similarity tolerance, which must be carefully set in order to sufficiently smooth the volume whilst preserving uncommon features, should be carefully investigated.

Ghosting artifacts were found to span several voxels, in contrast to the small scale of most noise, and caused false positives to form in the BG thin structure map. While these were ignored thanks to the purposeful over-estimation of the ventricles in the

mask created by the large CSF region removal procedure (Section 3.4.7), this artifact warrants correction since it is not guaranteed to be smoothed out and thus eliminated during resolution enhancement; it was subsequently a prime cause of mislabelling WM at GM/CSF boundaries in Figs 5.10 and 5.15. If the blood vessel network in B^{BG} is given a structural model as advocated in Section 6.4.1, then the ghosts can be identified by exploiting their lack of connectivity to the rest of the network; a further constraint could enforce that ghosts run close and roughly parallel to voxels flagged in B^{CSF} (as pictured in Fig 6.7.

Since the tissue type of the affected voxels cannot be simply and reliably determined, a function of their neighbouring voxel intensities can replace them, as advocated by the outlier-rejecting smoothing approach of You and Crebbin [You and Crebbin, 1995]. In Fig 6.8 the median of the immediate neighbours – including the hyperintense CSF, hypointense ghosts and the surrounding brain tissue – in a $3 \times 3 \times 3$ cube replaced putative ghost voxels (marked with aqua arrows) and led to a reduction in their intensity extrema without noticeably affecting the location of the ventricle borders. The localisation of the affected voxels could be achieved by the intersection of areas running close and parallel to the borders of B^{CSF} with regions in B^{BG} not connected to the rest of the vessel network (as pictured in Fig 6.7). Nevertheless, until the precise cause of these artifacts has been determined then it may not be safe to make such corrections.

6.5 Further applications

Aside from the common applications of class content maps listed in Section 1.3.1, the brain tissue labelling produced in this thesis may now also be applied to several other fields for which the accuracy at small scales or the precise quantification of PVE may be crucial. These include assistance for multi-tensor model disambiguation in DTI (Section 6.5.1), automatic placement of ROIs covering blood vessels (Section 6.5.2), identification of homogeneous tissue regions to better estimate bias fields (Section 6.5.3), as well as guidance of multi-modal (or multiple image weighting factor) volume co-registration through the congruence of anatomical features (Section 6.5.4).

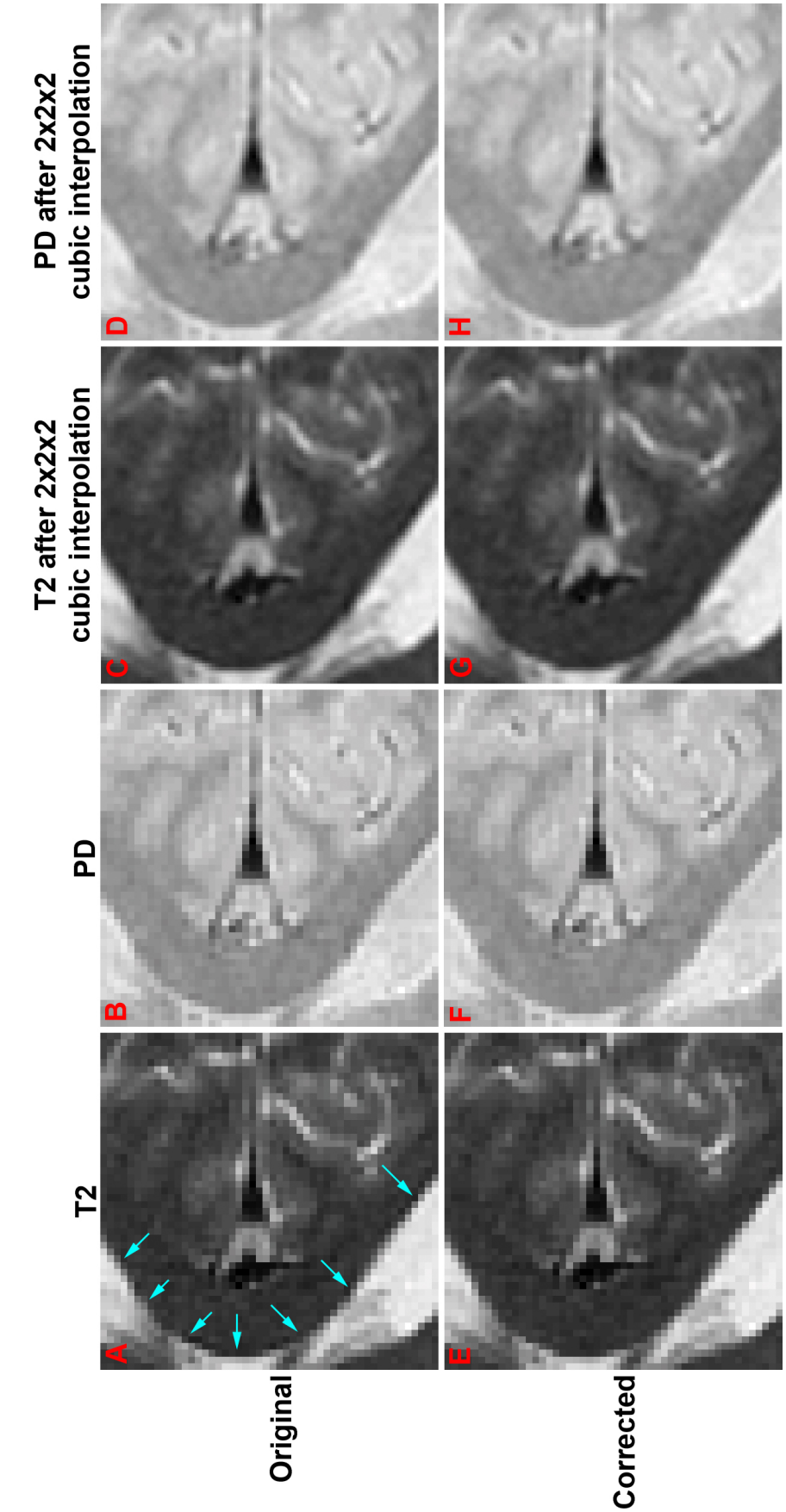

Figure 6.8: **Correction of ghosting**. **A** & **B**) Example of ghosting reproduced from Fig 3.17D. T_2 and PD data shown. **C** & **D**) **A** and **B** after cubic interpolation, upscaled by a $2 \times 2 \times 2$ ratio. **E** & **F**) **A** and **B** after correction of vessel ghosting by median filtering of the putative affected voxels. **G** & **H**) **E** and **F** after cubic interpolation, upscaled by a $2 \times 2 \times 2$ ratio. Aqua arrows are referenced in Section 6.4.2.

6.5.1 Prior information for DTI tractography

Tractography is a non-invasive method for investigating WM fibre connectivity using DTI (described in Section 1.1.5), which has been acquired in the LBC1936 dataset to quantify some effects of ageing [Bastin et al., 2009]. In WM regions the extracted water diffusion profiles in the centre of large tracts are highly anisotropic, in contrast to GM areas where they are less so and also to CSF-filled regions which have highly isotropic profiles. The assumption that a unique and accurate main diffusion direction (defined by the eigenvector corresponding to the dominant eigenvalue of a single diffusion tensor) can be prescribed at each voxel is also affected by SNR, motion, other artifacts and sequence parameters [Lazar and Alexander, 2003][Lazar and Alexander, 2005][Lazar et al., 2005][Anderson, 2001]. PVE is applicable to the formation of diffusion tensors where WM fibres within a voxel cross, split or merge; it is also relevant where WM terminates in GM or borders the ventricles, for which prior knowledge of the spatial localisation of those regions (obtained from tissue labelling) may prove useful.

Confounding the direct application of tissue content maps are observations that the results of fibre-tracking algorithms often proceed directly through ventricles, sulci and blood vessels (for example in [Jbabdi et al., 2007]) identified on an overlaid structural volume. While the tract progression and termination criteria may be to blame, DTI is acquired using echo planar imaging (EPI) and hence its non-rigid distortions may cause displacement of tracts into non-WM structures. If registration error could be eliminated then two useful priors could be formed using the tissue labels generated in Chapter 5.

The first is an indicator for whether a one- or two-tensor diffusion model should be selected, depending on the number of tissue class constituents in each voxel's label (which may need to be downsampled to reach the DTI volume's resolution). Like the constrained mixel dimension described in Section 5.1, it is assumed that up to two different diffusion profiles can mix inside a voxel and mixing models (i.e. [Alexander et al., 2001]) have been proposed for this purpose. The likely shape of each diffusion profile can also be generated: WM will be a highly anisotropic ellipse, CSF will be larger and isotropic, and GM will lie somewhere in between. While high-angular resolution diffusion imaging [Tuch et al., 2002] and Q-space imaging [Tuch et al., 2003][Tuch, 2004] methods can already produce a two-tensor orientation distribution function, the imaging time required limits their application in clinical settings and so

this prior will be most beneficial to methods attempting to fit two-tensor models in DTI acquisitions that typically produce single tensors.

Secondly, FA-based thresholds for tract progression and termination can be supplemented with WM-content-based probabilities of tract membership (in the case of fuzzy [Sotiropoulos et al., 2008] and probabilistic [Behrens et al., 2003] tractography methods) or high path costs for voxels not containing WM (for linear tract-following approaches such as [Mori et al., 1999]). In the case of single-tensor imaging these changes may allow FA thresholds to be made less strict so that more voxels containing crossing fibres can contribute to the results. Since fibres pass through deep GM nuclei, which were labelled as pure GM by the expert and the proposed methods often in Section 5.4.6, then label-based priors in these regions may not be as informative.

6.5.2 Automated placement of territorial arterial input functions for perfusion MRI

Dynamic contrast enhanced T_1-weighted MRI allows the inspection of blood perfusion properties inside the parenchyma using an invasive contrast agent. The concentration-time curve which is extracted must be deconvolved using an arterial input function (AIF) which describes changes in the agent's concentration inside blood vessels over time. AIFs are typically estimated from the mean signal intensity of a major artery, and are usually defined using circular ROIs placed manually over the cross-section of vessels proceeding through axial slices (as shown in Fig 6.9). High specificity of the BG class labelling (combined with structural modelling of the vessels) presents the possibility for irregularly-shaped ROIs to be sited at least semi-automatically. It is expected that manual pruning would be needed to remove unsuitable vessels such as large veins – making precise discrimination dependent on their relative position to other vessels as well as to their size and shape, and thus difficult to fully automate – and so this approach can be appreciated as a labour-saving method for generating AIF ROI candidates.

It has been noted that the relevance of the AIF is reduced for tissues distal to its particular ROI, and therefore numerous ROIs (territorial AIFs [Christensen et al., 2004]) can be placed throughout the volume. Vessels could be separated from other BG structures by their tubular shape – using a measure such as the plate-tube shape discriminant R^A (Eqn 3.7) if this could be made viable, as seen at high resolutions in Fig 3.19 – or by their connectivity to the rest of the circulatory network, requiring structural mod-

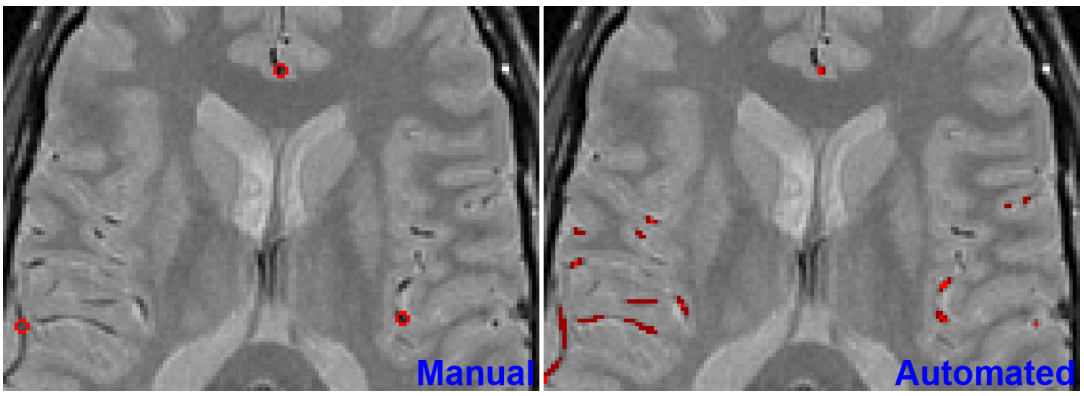

Figure 6.9: **Placement of territorial AIFs for perfusion imaging**. **Left**: Axial slice from a PD-weighted volume with circular ROIs (borders marked in red) placed manually over arteries. **Right**: Partial volume segmentation of some vessels with some candidate AIF regions (pure voxels highlighted in red, partial volumes in dark red) identified.

elling of the vessels (as investigated in Section 6.4.1). Since at clinical resolutions the vast majority of vessel voxels will be affected by PVE due to their small radii, then either only close-to-pure voxels should be selected (by a threshold on fractional BG content resulting in small AIF ROIs) or the PVE can accounted for by utilising the soft segmentation result or by exploiting the constant contribution of non-blood tissue to the changing MR signal [van Osch et al., 2001].

6.5.3 Bias correction through analysis of pure tissue uniformity

Intensity non-uniformity artifacts (also known as the bias field, described in Section 1.3.3) were not corrected in pre-processing (outlined in Section 2.2) due to poor performance of the BFC [Shattuck et al., 2001] and N3 [Sled et al., 1998] tools in large WM regions. Since the true spatial distributions of intensity can only be estimated – rather than precisely quantified – using prospective scanner calibration techniques for real brain images then retrospective methods are required (summarised in [Belaroussi et al., 2006]). A popular retrospective approach [Wells et al., 1996] (later used in FAST [Zhang et al., 2001]) corrected for the bias field by accounting for discrepancies between the original volume intensities with their reconstruction from class intensity distribution parameters and tissue labelling.

When this difference is formed using hard labelling, assumptions are made that only pure tissue voxels exist and that they have been correctly labelled. However, in the deep GM nuclei, at tissue boundaries where some voxels will be affected by PVE,

and especially where there is high spatial variation in the labelling such as temporal cortex, the falx cerebri, and the cerebellum then the validity of these assumptions break down. There is still potential for errors to occur at tissue boundaries using soft segmentation (as observed in Fig 5.16C and Fig 5.17C) due to the chance for label confusion and because high-frequency noise has not been accurately smoothed at edges. The following suggestion to improve bias correction methods is not concerned with the details of modelling the bias field, but instead with the sampling of volume labels and intensities away from boundaries in order to obtain better local bias estimates.

A simple sampling strategy would involve checking a label uniformity window at each voxel i over a neighbourhood Nh in which the labels F must be (at least almost) the same in order to be considered valid samples V to construct an estimate of the bias field (Eqn 6.1). $0 \leq \varepsilon \ll 2$ implements the uniformity tolerance of the local labelling, where 2 is the maximum absolute difference between two labels.

$$i \in V \text{ if } \forall j \in Nh(i), \left| f_i - f_j \right| < \varepsilon \tag{6.1}$$

In regions where the number of samples is low then this method will be more sensitive to labelling errors and pure tissue variability (as observed at the parenchymal borders in Fig 6.10C). There are several options for reducing this sensitivity: the size of Nh can be reduced, ε can be increased, or the universal quantifier in Eqn 6.1 can be relaxed to an existential check on a fraction of the neighbours – with a corresponding decrease in the confidence of the bias field estimate. On the other hand, when there are many samples available then the sampling could be restricted to pure tissue so any issues arising from the discretisation of the soft label set will be eliminated. Pure tissue sampling has been used in several other bias correction approaches (i.e. [Dawant et al., 1993]) though only for pure WM regions, due to their high expected intensity uniformity and lack of PVE. However, valuable information can also be gained from GM which was reduced in contrast in occipital regions (as observed in Fig 5.11). CSF may need to be excluded from the sampling due to the high variability in its pure tissue intensity; the effects can be seen in in Fig 6.10C, where it causes large deviations at the edges of the parenchyma.

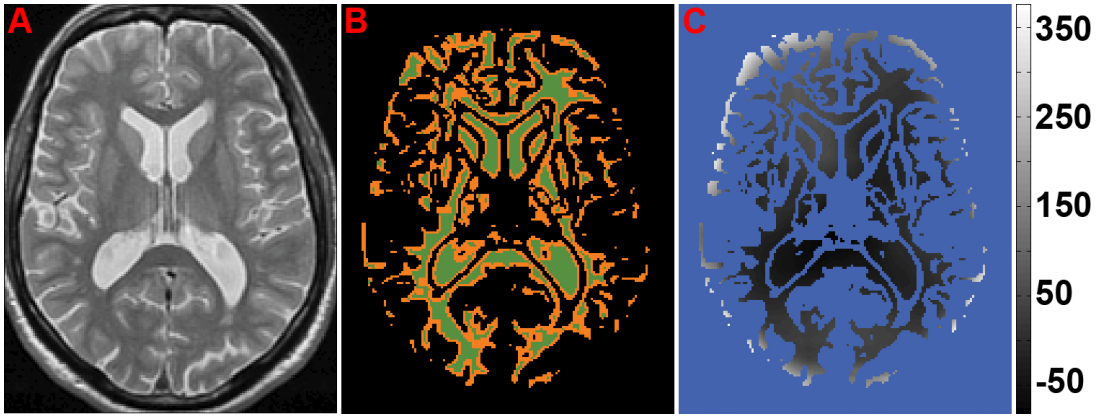

Figure 6.10: **Proposed voxel sampling method for bias field estimation**. **A**) PD data from the slice on which ROI 5 is located. **B**) After segmentation was performed using the parameters specified in Section 5.4 then universal uniformity of labelling in the local neighbourhood (highlighted in green) and existential uniformity over at least 19 similarly-labelled neighbours (highlighted in orange) were flagged using Eqn 6.1, with the low tolerance uniformity window label difference $\varepsilon = 0.2$. **C**) Deviation of the original intensities in **A** from the intensities reconstructed from the linear combination of the class content proportions and class means (obtained from segmentation using the parameter values listed in Section 5.4), following the mixel model (Dfn 5.2.1). This bias field estimate was smoothed by a Gaussian filter with $\sigma = 2$, using the uniformity mask defined in **B** (in blue), and plotted with respect to the gray-level scale to the right.

6.5.4 Multi-modal (or multiple image weighting factor) co-registration using label-based similarity

Dual-echo T_2-PD volumes are acquired simultaneously and therefore do not require co-registration (first examined in Section 1.3.2), but registration is necessary for the fusion of volumes for other image weighting factors or other modalities acquired of the same tissue (as suggested in Section 6.2) as well as for the transfer of labels or structure boundaries from a brain atlas (advocated in Section 6.4.1). The underlying issue being solved by registration is that if the same spatial locality is not being compared at each voxel then multi-channel segmentation, gradient calculation or feature transfer will not be relevant and lead to inaccuracies in processing. While this approach is suitable for multi-channel structural MRI acquisitions of a single subject, a labelling-based similarity metric will not viable when registration is desired to traditional non-adaptive atlases (in contrast to those proposed in Section 6.4.1) nor to volumes from other subjects since the assumption that exactly the same tissue exists in the same locations in each volume will be violated. Furthermore, the provision of anatomical context to fMRI or DTI volumes from structural data using this method must be tackled very carefully since there are distortions in intensity and shape as well as poor contrast, high levels of noise and worse resolution caused by the trade-offs involved in the rapid EPI acquisition process.

Similarity between selected features during transformation determines the quality of the matching in registration, and it has been noted that features such as edges and gradients may not be consistent across different modalities or image weighting factors, and they do not tend to scale well with increasing noise [Lester and Arridge, 1999]. However, the tissue labelling obtained from segmentation can be used as a matching feature [Zitova and Flusser, 2003] since the tissues generating the observed intensities are the same in each voxel of co-registered volumes, and thus the labelling should be identical once the volumes are aligned. [Rogelj and Kovacic, 2001] found that a segmentation-based similarity measure (based upon matching high probabilities of hard class labels) was more accurate than one based on local joint entropy of the intensities. The latter measure treated each intensity pair individually and could not consider any relations between intensities within the same tissue class.

The advantages of using the labelling as a feature are two-fold. First, simple metrics based upon the label difference can be employed to quantify the similarity. Secondly, the application of both appropriate class intensity models as well as processing

which promotes regional coherence and appreciation of structure shape during segmentation can make the labelling robust to imaging noise and poor contrast between tissue types. Individual modalities or image weighting factors may not contrast certain pairs of tissue classes well, such as BG and CSF in T_1 (observed in Fig 1.11), but the integration of multi-channel data segmentation and registration in [Wyatt and Noble, 2003] tackled this problem by moving from less complicated independent estimation of each channel's segmentation parameters to a more complex simultaneous estimation process once registration had improved. The advantage of this strategy, akin to the switch between diffusion-based resolution enhancement and structural-model-guided reconstruction suggested in the introduction to this chapter, is that the use of somewhat-aligned multi-channel data would likely cause fewer local optima to form in the search space as well as increase the capture range and distinctiveness of the global optimum. In the initial stage then the poorly-contrasted classes may need to be combined together and the the joint class histogram can be investigated to determine similarity.

Once the features to match have been finalised, then the optimisation of the similarity metric could involve the hierarchical correction of large global differences (such as affine transformations to align the parenchymal masks and large regions of contrasting tissue such as the ventricles) first before local small-scale registration errors (though non-rigid transformations to match, for example, the cortex and its sulci) could be tackled. At this scale high spatial variation of the labelling will be useful to drive transformation based on derivatives of the similarity measure, and so the performance of the proposed segmentation methods at tissue boundaries (particularly for thin structures) will need to be excellent.

6.6 Publications

The algorithms described in this thesis have been implemented using both multi-platform MATLAB code and compilable C-based MEX files, and will be made available in the RESCUE segmentation project at Sourceforge[5] after publication of this thesis. The project contains four main components:

- **TSD**: thin structure detection, implementing the methods detailed in Chapter 3, which were inspired by the work of [Frangi et al., 1998] and [Westin and

[5] At `http://sourceforge.net/projects/rescueseg`.

Knutsson, 2003];

- **VRE**: volume resolution enhancement, implementing the methods detailed in Chapter 4, which were inspired by the work of [Salvado et al., 2006] and [Salvado and Wilson, 2006];
- **FCM**: FCM segmentation, which was inspired by the work of [Ahmed et al., 2002];
- **SEG**: the main segmentation algorithm, implementing the methods detailed in Chapter 5, which were inspired by the work of [van Leemput et al., 2003] and [Li et al., 1997].

All the source code will be made free to download and modify within restrictions of the GNU General Public License, which provides opportunities for other authors to contribute towards further development and refinement. MRI data import methods can be found in the external *Tools for NIfTI and ANALYZE image* package[6] and therefore both plotting and reporting functions will also not be included in the publically-available code. The software tool designed to assist expert soft manual labelling (described in Section 2.4) as well as thin structure measurement (a form of parameter training detailed in Section 3.4.6) will be released separately in the future.

A basic version of the thin structure detection and volume resolution enhancement methods applied to FCM clustering was published in the proceedings of the Medical Image Understanding and Analysis conference 2008 [Withers et al., 2008]. This paper is reproduced in Appendix 7.3. A more detailed paper describing all the methods proposed in this thesis is currently in production.

6.7 Summary

This chapter contains the final evaluation of the methods proposed in this thesis, firstly (Section 6.1) in terms of their performance by the key metrics identified in Section 1.3.5 and their fulfilment of the original objectives listed in Section 1.4, and also in the future potential for improvement (Section 6.4) and also development in other fields (Section 6.5). The final part of this chapter (Section 6.6) details the written and software publications resulting from the work completed in this thesis.

[6]Written by Jimmy Shen and available at `http://www.mathworks.co.uk/matlabcentral/fileexchange/8797`.

Since the labelling of the LBC1936 dataset produced for this thesis, nor the range of testing performed, were not entirely satisfactory then these issues are further discussed in Sections 6.2 and 6.3. Of particular note in the latter section are proposals for soft labelling analysis and parameter training which may improve human operator performance.

Chapter 7

Appendices

7.1 Volume resolution enhancement flow tables

Orientation	**Direction**	**Distance**^{2}
X^+	east	1
Y^+	north	1
Z^+	up	h^2
X^+Y^+	north-east	2
X^-Y^+	north-west	2
X^+Z^+	up-east	$1+h^2$
X^-Z^+	up-west	$1+h^2$
Y^+Z^+	up-north	$1+h^2$
Y^-Z^+	up-south	$1+h^2$
$X^+Y^+Z^+$	up-north-east	$2+h^2$
$X^-Y^+Z^+$	up-north-west	$2+h^2$
$X^+Y^-Z^+$	up-south-east	$2+h^2$
$X^-Y^-Z^+$	up-south-west	$2+h^2$

Table 7.1: **Flow orientations, direction sets and relative distance weights in 3D**. The first three rows are the diffusion orientations along the data axes to the 1st order neighbours. The next 6 rows define flow to and from the 2nd-order neighbours, and the final four rows list the third-order neighbours which are adjacent to the central voxel. The **Direction** column denotes the directions of flow, along the prescribed orientation in the first column. The **Distance** column defines the squared Euclidean distance from a central voxel to the next voxel along the prescribed orientation, with a unit distance assigned to the axial-plane first-order neighbours and h (the ratio between the slice thickness and the axial-plane voxel dimensions) assigned to the out-of-plane first-order neighbours.

L	**1^{st}-order**	**2^{nd}-order**	**3^{rd}-order**	**max $\Delta\tau$**
$1\times1\times1$	0	0	0	1
$1\times1\times2$	1	0	0	$\frac{1}{1+1}$
$1\times1\times(\geq3)$	2	0	0	$\frac{1}{1+1+\frac{1}{h^2}}$
$1\times2\times1$	1	0	0	$\frac{1}{1+1}$
$1\times2\times2$	2	1	0	$\frac{1}{1+1+\frac{1}{h^2}+\frac{1}{1+h^2}}$
$1\times2\times(\geq3)$	3	2	0	$\frac{1}{1+1+\frac{2}{h^2}+\frac{2}{1+h^2}}$
$1\times(\geq3)\times1$	2	0	0	$\frac{1}{1+2}$
$1\times(\geq3)\times2$	3	2	0	$\frac{1}{1+2+\frac{1}{h^2}+\frac{2}{1+h^2}}$
$1\times(\geq3)\times(\geq3)$	4	4	0	$\frac{1}{1+2+\frac{2}{h^2}+\frac{4}{1+h^2}}$
$2\times2\times1$	2	1	0	$\frac{1}{1+2+\frac{1}{2}}$
$2\times2\times2$	3	3	1	$\frac{1}{1+2+\frac{1}{2}+\frac{1}{h^2}+\frac{2}{1+h^2}+\frac{1}{2+h^2}}$
$2\times2\times(\geq3)$	4	5	2	$\frac{1}{1+2+\frac{1}{2}+\frac{2}{h^2}+\frac{4}{1+h^2}+\frac{2}{2+h^2}}$
$2\times(\geq3)\times1$	3	2	0	$\frac{1}{1+3+\frac{2}{2}}$
$2\times(\geq3)\times2$	4	5	2	$\frac{1}{1+3+\frac{2}{2}+\frac{1}{h^2}+\frac{3}{1+h^2}+\frac{2}{2+h^2}}$
$2\times(\geq3)\times(\geq3)$	5	8	4	$\frac{1}{1+3+\frac{2}{2}+\frac{2}{h^2}+\frac{6}{1+h^2}+\frac{4}{2+h^2}}$
$(\geq3)\times(\geq3)\times1$	4	4	0	$\frac{1}{1+4+\frac{4}{2}}$
$(\geq3)\times(\geq3)\times2$	5	8	4	$\frac{1}{1+4+\frac{4}{2}+\frac{1}{h^2}+\frac{4}{1+h^2}+\frac{4}{2+h^2}}$
$(\geq3)\times(\geq3)\times(\geq3)$	6	12	8	$\frac{1}{1+4+\frac{4}{2}+\frac{2}{h^2}+\frac{8}{1+h^2}+\frac{8}{2+h^2}}$

Table 7.2: **Number of neighbours to which flow is not constrained during the reverse diffusion stage given different upscaling ratios**. The ordering of the first and second dimensions is not deemed crucial since they are assumed to possess the same voxel dimensions, whereas the third through-slice dimension has a thickness h compared to unit axial-plane dimensions. **L**: ratio for upscaling the original volume. **Central columns**: number of sub-voxels in certain orders of the neighbourhood which are made accessible by the lack of voxel boundary flow constraints. **max** $\Delta\tau$: maximum allowed value of the timestep with respect to the CFL bound (Eqn 4.4). The distance weights are the inverse square of the Euclidean distance from the central voxel.

7.2 FAST results with bias correction enabled

	Absolute error					*κ-statistic score*					*Volume error*			
ROI	**BG**	**WM**	**GM**	**CSF**	**All**	**BG**	**WM**	**GM**	**CSF**	**All**	**BG**	**WM**	**GM**	**CSF**
1	10	228	318.5	146.5	703	0.372	0.702	0.562	0.749	0.643	8	-98	9.5	80.5
2	33	391	559.5	269.5	1253	0.000	0.473	0.286	0.492	0.357	33	163	-217.5	21.5
3	42	417.5	633.5	345	1438	0.207	0.319	0.230	0.446	0.272	-8	-30.5	-212.5	251
4	0	547.5	843	435.5	1826	0.000	0.179	0.062	0.423	0.167	0	-339.5	370	-30.5
5	18	302.5	445	159.5	925	0.436	0.625	0.529	0.689	0.573	4	121.5	-232	106.5
6	6.5	648	706.5	106	1467	0.000	0.271	0.320	0.826	0.418	6.5	433	-476.5	37
7	54.5	473	457	99.5	1084	0.113	0.483	0.414	0.795	0.513	52.5	-420	360	7.5
8	3.5	453	491	47.5	995	0.299	0.313	0.364	0.774	0.388	1.5	116	-134	16.5
9	8.5	140	277.5	173	599	0.550	0.790	0.605	0.640	0.651	1.5	-23	-136.5	158
10	12.5	348.5	476.5	213.5	1051	0.776	0.562	0.371	0.510	0.429	8.5	276.5	-234.5	-50.5

Table 7.3: **Error statistics for FAST using bias field correction over expert-labelled ROIs from the LBC1936 dataset**. Errors are listed to 3 decimal places. The smallest absolute error and largest κ statistics between the proposed method (listed in Table 5.4) and FAST are highlighted in red; equal values are highlighted in blue.

7.3 Publication in MIUA 2008

Multi-scale segmentation of dual-channel MRI using volume resolution enhancement and tubular structure detection

James Withers[a*], Mark Bastin[b] and Amos Storkey[c]

[a]Neuroinformatics Doctoral Training Centre, University of Edinburgh, [b]Medical Physics, University of Edinburgh, [c]Institute for Adaptive and Neural Computation, University of Edinburgh

Abstract. Brain tissue segmentation is complicated by noise and partial volume voxels that contain mixtures of two or more tissue types. We examine three extensions to segmentation methods which may allow more accurate classification of small-scale structures while still providing robustness to noise. Firstly, the location and orientation of fine tubular structures can be identified using differential geometry. Subsequent neighbourhood filtering operations, for ensuring local homogeneity or performing resolution enhancement of the volume, may then operate with different scales and orientations according to the structure of interest. Finally, dual-channel T2-PD segmentation can be improved by weighting the information from each modality to better emphasise contrasting tissues. We present quantitative results on simulated volumes and qualitative assessments on a real dataset, discuss limitations and improvements to the methods, and examine future application to an Expectation-Maximisation framework.

1 Introduction

The automatic segmentation of brain volumes – usually into gray matter, white matter, cerebrospinal fluid (CSF) and blood/background – is necessary for analyses based on tissue type and volume quantification. MRI is able to produce a range of tissue contrasts, but in clinical practice confounds to successful segmentation include various sources of noise, low resolution, ambiguous partial volumes (where multiple tissues reside in a single voxel), and poor contrast between some tissues in certain modalities. Many approaches to limiting the effects of noise enforce local class or intensity homogeneity (i.e. [1]). An unfortunate consequence of processing neighbourhood interactions at a single scale is that small details can be smoothed over like the noise. In particular, Diffusion Tensor Imaging can benefit from accurate and detailed tissue maps in order to provide robust procession and termination criteria for tractography methods.

In this paper we present several extensions to previously-developed methods that aid segmentation; primarily, we incorporate multi-scale oriented filters for protecting thin structures while still providing robustness to noise, introduced by Svensson *et al*. [2]. Brain atlases (as used in [3]) are formed from averages over a population and thus cannot provide strong priors for the precise layouts of cortical sulci and blood vessels which are unique between brains. Sato *et al*. [4] have investigated using differential geometry in general multi-channel segmentation, and Descoteaux *et al*. [5] have used similar filters to detect the centrelines of blood vessels. This approach is generally fast and robust, but morphological methods (such as [6]) have also been used to detect such structures. The majority of segmentation approaches deal with T1-weighted volumes due to their excellent gray and white matter contrast, but the best modalities for detecting cortical sulci and blood vessels were found to be T2 and proton density (PD) respectively. Consequently steps are taken to exploit multi-channel data in segmentation, as investigated by Liang *et al*. [7]. Finally, in order to process small structures at more relevant scales, volume upsizing with partial volume reduction [8] is employed. A further anisotropic smoothing step [9], first investigated in brain MRI by Gerig *et al*. [10], aims to reduce noise while also preserving detail.

2 Methods

The implementation of Fuzzy C-Means (FCM) clustering by Ahmed *et al*. [11] was chosen to demonstrate the following improvements. Given intensities X (indexed by voxels i) and centroids V for C classes, the objective function J weights the information gained from a central voxel with that from its $3 \times 3 \times 3$ neighbourhood N using parameter α:

$$J = \sum_{i \in volume} \sum_{k=1}^{C} u_{ik}^{p} \left\| x_i - v_k \right\|^2 + \frac{\alpha}{\text{numel}(N)} \sum_{i \in image} \sum_{m=1}^{C} u_{im}^{p} \sum_{j \in N_i} \left\| x_j - v_m \right\|^2 \quad (1)$$

In noisy images α should be increased to produce a more homogeneous result. u defines the tissue class memberships at each voxel, and p is a positive exponent controlling the fuzziness of the segmentation result. The algorithm is fast and robust, but tends to erode small details due to the drive for homogeneity using a single neighbourhood filter, and it can perform poorly when contrast is low – such as between background/blood and CSF in T1-weighted volumes.

*E-mail: s0456265@sms.ed.ac.uk

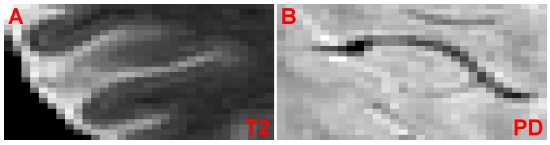

Figure 1. Thin structures in brain volumes. (**A**) Cortical sulci (bright in T2) can possess complex topology, often flaring near the brain surface. (**B**) Blood vessels (dark in PD) are tubular structures that form smooth arcs and junctions.

2.1 Thin structure detection using differential geometry

Tubular structures possess a direction of minimal intensity curvature which coincides with the orientation of the tube, and an orthogonal direction (or plane in 3D) of maximal curvature where intensity reaches a local extremum at the medial axis. Blood vessels are naturally tubular but thin cortical sulci also appear as such in 2D (Fig 1). Sulci are CSF-filled fissures with a tightly-packed and complex 3D structure caused by the folding of the cortex, making 3D filters that may respond optimally to plate-like shapes slightly less accurate than their 2D counterparts at the appropriate orientation. The proposed filter for locating both the presence and orientation of tubular structures is based on eigenvalue analysis of the Hessian [12]: the eigenvector associated with the smallest eigenvalue λ_2 represents the direction of minimal curvature, and the eigenvector of λ_1 gives the direction of maximal curvature. The sign of λ_1 discriminates between local intensity minima and maxima, so MRI modalities can be chosen that best contrast the structures of interest: CSF is much brighter than other tissues in T2-weighted volumes, and blood is very dark in PD volumes.

Frangi *et al*. [12] introduce R_B, which assesses adherence to an optimal tube profile ($\lambda_1 \gg \lambda_2$), and S, the Euclidean norm of the eigenvalues, which is small for noise and highest in blob-like areas, which include structures proceeding in the through-plane direction and junctions. A parameter s refers to the scale space level for the Hessian, which allows structures of different sizes to be optimally detected. The tubular structure filter T (Eqns 2 and 3) took the maximum response over a range of scales, but because tissue contrasts are dependent on the scan sequences used, a threshold was set to achieve a $\sim$99% positive predictive value on labelled data for voxels containing the mapped tissue.

$$T_{sulci}(s) = \begin{cases} 0, & \text{if } \lambda_1 \geq 0; \\ (1 - e^{-S}) * e^{-R_B}, & \text{if } \lambda_1 < 0. \end{cases} \quad (2)$$

$$T_{blood}(s) = \begin{cases} 0, & \text{if } \lambda_1 \leq 0; \\ (1 - e^{-S}) * e^{-R_B}, & \text{if } \lambda_1 > 0. \end{cases} \quad (3)$$

2.2 Volume resolution enhancement to reduce the relative amount of partial volume voxels

Salvado *et al*. [8] developed an interpolation scheme where intensity could diffuse between subvoxels on an upsized volume to reduce the relative amount of partial volume on tissue boundaries. This process, along with a further anisotropic smoothing step [9], will be referred to as *volume resolution enhancement*. Gaussian smoothing reduces both noise and the intensity extrema of thin structures when calculating the diffusion gradients, but the pre-computed minimum curvature direction is able to bias the filtering: a smaller, tighter, ellipsoid 3D Gaussian oriented in parallel to the structure can better preserve the intensity (Fig 2). Also, less diffusion occurs than expected around thin contrasting structures since local rank ordering limits (which dictate the amount of possible flow) can sometimes overlook their extreme intensties. The limits are set further from the median in these regions.

2.3 Modifications to Fuzzy C-Means clustering

Multi-scale oriented filters can also be applied to neighbourhood schemes for segmentation that aim to homogenise the local class or intensity. In our exemplar segmentation method (Eqn 1), the intensities of 26 neighbours are considered of equal importance. A larger neighbourhood can be considered for improved noise reduction by employing a normalised Gaussian filter G to weight the local information by the distance from the central voxel (*dist*) in a similar fashion to

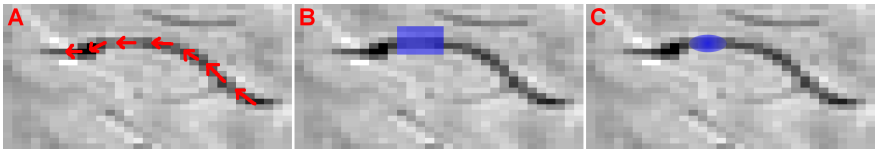

Figure 2. Multi-scale oriented filters help to preserve thin structures. (**A**) The direction of minimal curvature (red arrows) can be mapped using differential geometry. (**B**) Thin structures are blurred due to their small size compared to the isotropic filter dimensions (blue box). (**C**) By tailoring the filter size, profile and orientation (radial blue ellipse), the valid neighbourhood is altered and the structures can be better preserved.

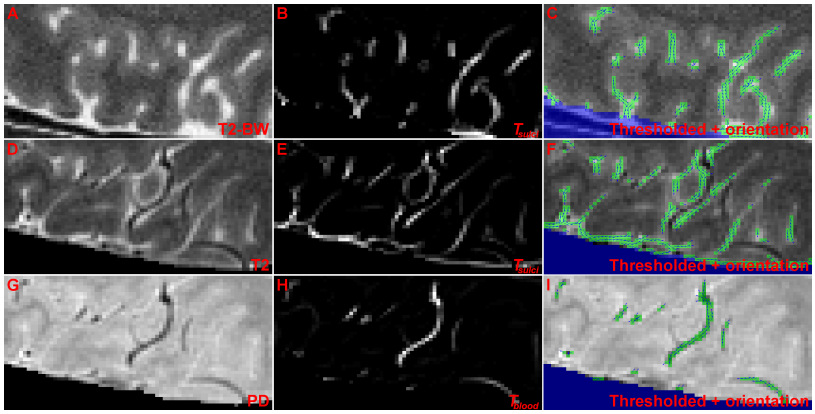

Figure 3. Detection of thin structures. (**A**-**C**) Detection of cortical sulci in temporal cortex of a BrainWeb T2-weighted volume. (**D**-**F**) Detection of sulci in temporal cortex of a real T2-weighted volume. (**G**-**I**) Detection of blood vessels in temporal cortex of a real PD volume. **Left**: volume intensities. **Middle**: Continuous map of T. **Right**: Thresholded T in green, non-brain voxels in blue, and direction of minimal curvature denoted with small blue arrows.

Shen *et al*. [13]. Thin tubular structures can be protected by orienting the filters and reducing their profile σ in these areas. The updated form of Eqn 1 uses resolution-enhanced T2-weighted and PD volumes $\bar{X}$ indexed by subvoxels z:

$$J_{modified} = \sum_{z \in volume} \sum_{k=1}^{C} u_{zk}^{p} \left\| w_k(\bar{x}_z - v_k) \right\|^2 + \alpha \sum_{z \in volume} \sum_{m=1}^{C} u_{zm}^{p} \sum_{j \in N_z} G(dist(z,j), \sigma_z) \left\| w_m(\bar{x}_j - v_m) \right\|^2 \quad (4)$$

The weight matrix W exploits the quality of modality-specific tissue class contrasts: it enforces that well-contrasting tissue in one modality is more sensitive to deviation from its centroid than poorly contrasting tissue. The tuning of W could be approached by searching to maximise performance over a labeled dataset, but the values were set intuitively in the following tests. T2 has a greater weight for its bright CSF but less for white matter, which is poorly separated from the remaining classes. PD has an increased weight for the well-contrasted white matter, but reduced for CSF, which can be difficult to distinguish from gray matter. The other weights were set to unity.

3 Results

5 central axial slices of synthetic brain volumes generated by BrainWeb [14] with 3% added noise were used for quantitative evaluation. The ground truth is already known since the volumes are constructed from a set of constituent class maps. Real T2-weighted and PD MRI volumes were acquired using a dual-echo Fast Spin Echo sequence on a 1.5T scanner at the Western General Hospital, Edinburgh, UK. The voxels were isotropic and other scan parameters were NEX = 1, TE_{PD} = 15ms, TE_{T2} = 100ms, FOV = 240mm and matrix size 256×256. As a preprocessing step the brain was extracted from the volume and BFC [15] was used to correct for the bias field.

3.1 Tubular structure detection accuracy

BrainWeb volumes do not contain blood vessels, and so quantitative evaluation was performed on cortical sulci only using $s \in [0.5...0.85]$, corresponding to the expected range of the shorter dimension of the structures. The evaluation criterion was the correct presence of any CSF in voxels flagged by a threshold on T_{sulci}; since the threshold was set to ensure a $\sim$99% positive predictive value, the other error statistics are more informative (Table 1). The frequency of false positives doubles at the 5% BrainWeb volume noise level using the same threshold, but this error rate is still extremely small. Furthermore, since the small structure maps are not being used as tissue class priors but instead only as detail filter selectors, errors will only manifest as increasing the influence from immediate (perhaps noisy) neighbours and ideally have minimal impact on the segmentation result. Only half of the voxels containing CSF are detected, but a significant proportion of the remainder are found in larger blob-like structures such as the ventricles. These areas do not display a thin profile and therefore should not be detected. Fig 3 shows the performance of T_{sulci} in real and simulated images, and T_{blood} in real images only. The same threshold was applied to both rounds of sulcal detection. Blob-like structures tend to scatter the directions of minimal curvature, as would be expected, whereas their continuity along thin sulci and blood vessels is generally excellent.

	Ground truth		
	CSF present	**No CSF present**	
Thresholded	12209 (TP)	192 (FP)	**PPV:** 98.45%
Not thresholded	12613 (FN)	75076 (TN)	**NPV:** 85.62%
	Sensitivity: 49.19%	**Specificity:** 99.74%	

Table 1. Error statistics of CSF voxels flagged by thin structure detection against the BrainWeb ground truth. 100090 voxels were examined in total. TP = true positive. FP = false positive. TN = true negative. FN = false negative. PPV = positive predictive value. NPV = negative predictive value.

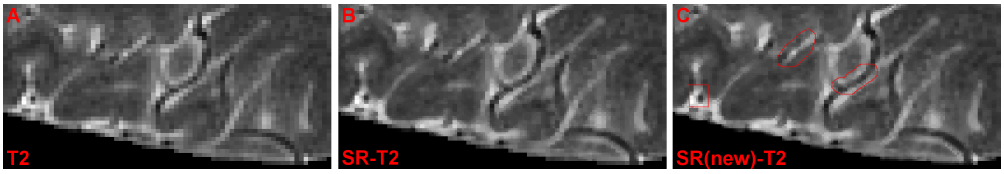

Figure 4. Improved crispness of resolution-enhanced thin structures. (**A**) Temporal cortex in a real T2-weighted volume. (**B**) Original $2 \times 2 \times 2$ resolution-enhanced version of (A). (**C**) Improved resolution enhancement of (A).

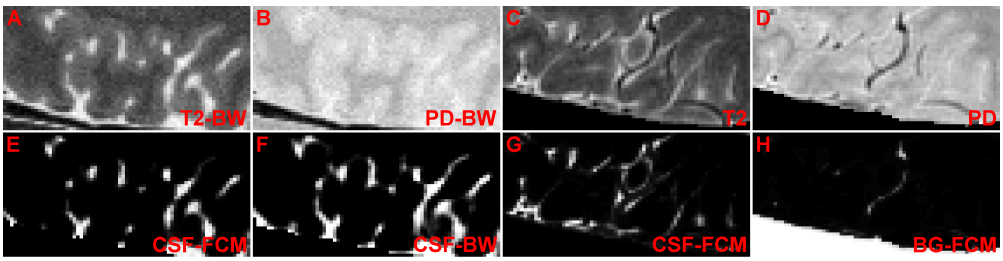

Figure 5. Segmentation accuracy. (**A** & **B**) Temporal cortex in BrainWeb T2-weighted and PD volumes. (**C** & **D**) Similar area in real volumes. (**E**) CSF memberships from (A) and (B) using modified FCM. (**F**) Ground truth CSF proportions. (**G** & **H**) CSF and blood/background memberships from (C) and (D) using modified FCM.

3.2 Changes in volume resolution enhancement quality

The results in Figs 4 and 5 and Table 2 were produced over 30 iterations with the diffusion speed $dt = 0.2$, using a $2 \times 2 \times 2$ resizing ratio. Resolution enhancement of a real T2-weighted image of temporal cortex is shown in Fig 4; the ringed areas show particular areas of improvement, in terms of the crispness of tissue borders.

3.3 Improvements in overall segmentation quality

The class centroids were initialised uniformly in the intensity space and clustering continued until the sum of their changes was less than 0.01. The modified process (Eqn 4) consistently outperformed the original (Eqn 1) in segmenting dual-channel data (Table 2). Since FCM produces class memberships, rather than physical proportions of mixtures, the difference between the output and the ground truth is displayed purely for comparative purposes. In real brain volumes the partial volume effects caused by the thinness of blood vessels mean they appear closest in intensity to other classes. As FCM is not able to intelligently disambiguate mixtures, the vessels in Fig 5H are mostly labelled as white matter, with darker regions gaining some contribution from blood, despite their excellent detection (Fig 3H).

4 Discussion

Our applications of eigenvalue analysis include increasing allowed flow during volume resolution enhancement processes, using the minimum curvature direction to change neighbourhood filter orientations, and tightening filter profiles in thresholded areas, in order to encourage the preservation of thin structures such as cortical sulci and blood vessels. We have presented a unified and simple model for detecting these structures which can achieve a good sensitivity while maintaining excellent specificity. T2-weighted and PD volumes produce the best contrast for these structures and by introducing a weighting vector to compensate for the individual modalities' poor contrasts and to exploit exceptional ones, then this dual-channel data is able to discriminate well between all four tissue types. T1-weighted volumes typically have poor contrast between CSF and background/blood, both appearing dark. T2-weighted and PD volumes can be acquired with a dual-echo sequence, without adding registration artefacts, enabling dual-channel segmentation.

The most costly parameter is the upsizing ratio for volume resolution enhancement: greater reduction in partial volume at tissue borders comes at the cost of increased the time and memory requirements, so practical concerns limit this ratio

	BrainWeb	**Original FCM**		**Modified FCM**	
	# voxels	Diff ($\pm\sigma$)	✓label	Diff ($\pm\sigma$)	✓label
WM	48732	0.12 ± 0.09	77.98%	0.11 ± 0.09	78.63%
GM	65230	0.13 ± 0.11	64.20%	0.10 ± 0.09	67.29%
CSF	25947	0.04 ± 0.06	49.56%	0.03 ± 0.08	51.87%
All	100090	0.30 ± 0.21	93.57%	0.26 ± 0.21	94.67%

Table 2. Evaluation of FCM clustering changes. *Diff*: mean difference between class memberships u and BrainWeb ground truth. ✓*label*: frequency of matching the maximum classes in u and ground truth. *# voxels*: number of voxels containing any of this tissue. WM = white matter. GM = gray matter.

to $3 \times 3 \times 3$. The regular and detail filter profiles (the standard deviation in each axis) were both proportional to this ratio. Large changes to the rank order limits cause adverse effects on the robustness of the resolution enhancement, so the change in detail areas was kept small. If the distribution of the widths of sulci or vessels change, due to age-related effects or altering the imaging resolution, a different range of Hessian analysis scales may be necessary.

Most sulci in the BrainWeb slices studied proceeded through the axial plane and only this 2D plane was considered in the production of T; by selecting the maximum response over several planes (i.e. sagittal and coronal) the sensitivity should improve and the minimal curvature vector should be more realistic. Alternatively, further investigation of 3D structure measures discussed by Frangi *et al.* [12] may prove fruitful, and the detection sensitivity of blood vessels – often badly affected by partial volume – is expected to increase with 3D processing due to their cylindical profile. Furthermore, their maps could benefit from diffusion processes aiming to reduce noise along the vessels' lengths [16]. We have shown the intensity curvature to extrema that gives the optimum responses to T is applicable in real volumes but the resulting orientation field could be made more robust by using a filtering step [17]. In order to further tailor the structure preservation process the filter profiles could be changed proportionally to the optimal scale for each voxel.

FCM has been shown to perform poorly in the disambiguation of tissue mixtures; the proposed application of this work is to improve more advanced tools. *FAST* [1] uses a Markov random field prior to influence the labeling of voxels, that could benefit from employing multi-scale oriented filters to preserve thin details. *Segment* [3] uses prior class maps but only weak priors can exist for the uniquely-positioned blood vessels and cortical sulci. T_{sulci} and T_{blood} could provide stronger priors to help disambiguate the mixture components, if they maintain high positive predictive values.

Acknowledgements

This work is funded by the Engineering and Physical Sciences Research Council and the Medical Research Council. We are grateful to Olivier Salvado for supplying the volume resolution enhancement code.

References

1. Y. Zhang, M. Brady & S. Smith. "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm." *IEEE Transactions on Medical Imaging* **20(1)**, pp. 45–57, 2001.
2. B. Svensson, M. Andersson, O. Smedby et al. "Efficient 3-D adaptive filtering for medical image enhancement." In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 996–999. 2006.
3. J. Ashburner & K. Friston. "Unified segmentation." *Neuroimage* **26(3)**, pp. 839–851, 2005.
4. Y. Sato, C.-F. Westin, A. Bhalerao et al. "Tissue classification based on 3D local intensity structures for volume rendering." *IEEE Transactions on Visualization and Computer Graphics* **6(2)**, pp. 160–180, 2000.
5. M. Descoteaux, D. Collins & K. Siddiqi. "Geometric flows for segmenting vasculature in MRI: theory and validation." In *Proceedings of Medical Image Computing and Computer-Assisted Interventation*, pp. 500–507. 2004.
6. L. Lorigo, O. Faugeras, W. Grimson et al. "CURVES: Curve evolution for vessel segmentation." *Medical Image Analysis* **5**, pp. 195–206, 2001.
7. Z. Liang, J. MacFall & D. Harrington. "Parameter estimation and tissue segmentation from multispectral MR images." *IEEE Transactions on Medical Imaging* **13(3)**, pp. 441–449, 1994.
8. O. Salvado, C. Hillenbrand & D. Wilson. "Partial volume reduction by interpolation with reverse diffusion." *International Journal of Biomedical Imaging* **1**, pp. 1–13, 2006.
9. O. Salvado & D. Wilson. "Thick slice interpolation using reverse anisotropic diffusion to reduce partial volume effect." In *Proceedings of IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pp. 1000–1003. 2006.
10. G. Gerig, O. Kubler, R. Kikinis et al. "Nonlinear anisotropic filtering of MRI data." *IEEE Transactions on Medical Imaging* **1(2)**, pp. 221–232, 1992.
11. M. Ahmed, S. Yamany, N. Mohamed et al. "A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data." *IEEE Transactions on Medical Imaging* **21(3)**, pp. 193–199, 2002.
12. A. Frangi, W. Niessen, K. Vincken et al. "Multiscale vessel enhancement filtering." In *Proceedings of Medical Image Computing and Computer-Assisted Interventation*, pp. 130–137. 1998.
13. S. Shen, W. Sandham, M. Granat et al. "MRI fuzzy segmentation of brain tissue using neighbourhood attraction with neural network optimization." *IEEE Transactions on Information Technology in Biomedicine* **9(3)**, pp. 459–467, 2005.
14. C. Cocosco, V. Kollokian, R.-S. Kwan et al. "Brainweb: online interface to a 3D MRI simulated brain database." *NeuroImage* **5(4)**, pp. S425, 1997.
15. D. Shattuck, S. Sandor-Leahy, K. Schaper et al. "Magnetic resonance image tissue classification using a partial volume model." *Neuroimage* **13**, pp. 856–876, 2001.
16. R. Manniesing, M. Viergever & W. Niessen. "Vessel enhancing diffusion A scale space representation of vessel structures." *Medical Image Analysis* **10(6)**, pp. 815–825, 2006.
17. C.-F. Westin & H. Knutsson. "Tensor field regularization using normalized convolution." In *Proceedings of the International Conference on Computer Aided Systems Theory*, pp. 564–572. 2003.

Bibliography

Aboutanos, G., Nikanne, J., Watkins, N., and Dawant, B. (1999). Model creation and deformation for the automatic segmentation of the brain in MR images. *IEEE Transactions on Biomedical Engineering*, 46(11):213–226.

Adelson, E. (2000). Lightness perception and lightness illusions. In *The New Cognitive Neurosciences*, pages 339–351.

Ahmed, M., Yamany, S., Mohamed, N., Farah, A., and Moriarty, T. (2002). A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 21(3):193–199.

Alexander, A., Hasan, K., Lazar, M., Tsuruda, J., and Parker, D. (2001). Analysis of partial volume effects in diffusion-tensor MRI. *Magnetic Resonance in Medicine*, 45:770–780.

Anbeek, P., Vincken, K., van Bochove, G., van Osch, M., and van der Grond, J. (2005). Probabilistic segmentation of brain tissue in MR imaging. *NeuroImage*, 27(4):795–804.

Anderson, A. (2001). Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magnetic Resonance in Medicine*, 46:1174–1188.

Antiga, L. and Steinman, D. (2004). Robust and objective decomposition and mapping of bifurcating vessels. *IEEE Transactions on Medical Imaging*, 23(6):704–713.

Arnold, J., Liow, J., Schaper, K., Stern, J., Sled, J., Shattuck, D., Worth, A., Cohen, M., Leahy, R., Mazziotta, J., and Rottenberg, D. (2001). Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage*, 13(5):931–943.

Ashburner, J. and Friston, K. (2005). Unified segmentation. *NeuroImage*, 26(3):839–851.

Aubert-Broche, B., Evans, A., and Collins, L. (2006a). A new improved version of the realistic digital brain phantom. *NeuroImage*, 32(1):138–145.

Aubert-Broche, B., Griffin, M., Pike, G., Evans, A., and Collins, D. (1997). Estimating the bias field of MR images. *IEEE Transactions on Medical Imaging*, 16:238–251.

Aubert-Broche, B., Griffin, M., Pike, G., Evans, A., and Collins, D. (2006b). Twenty new digital brain phantoms for creation of validation image data bases. *IEEE Transactions on Medical Imaging*, 25(11):1410–1416.

Avants, B. and Gee, J. (2003). The shape operator for differential analysis of images. In *Proceedings of International Conference on Information Processing in Medical Imaging*, pages 101–113.

Awate, S. and Whitaker, R. (2007). Feature-preserving MRI denoising: A nonparametric empirical Bayes approach. *IEEE Transactions on Medical Imaging*, 26(9):1242–1255.

Aylward, S. and Bullitt, E. (2002). Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE Transactions on Medical Imaging*, 21(2):61–75.

Azevedo, F., Carvalho, L., Grinberg, L., Farfel, J., Ferretti, R., Leite, R., Filho, W., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541.

Babalola, K., Cootes, T., Twining, C., Petrovic, V., and Taylor, C. (2008). 3D brain segmentation using active appearance models and local regressors. *Lecture Notes in Computer Science*, 5241:401–408.

Baker, S. and Kanade, T. (2002). Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1167–1183.

Barash, D., Israeli, M., and Kimmel, R. (2001). An accurate operator splitting scheme for nonlinear diffusion filtering. *Lecture Notes in Computer Science*, 2106:281–289.

Bastin, M., Clayden, J., Pattie, A., Gerrish, I., Wardlaw, J., and Deary, I. (2009). Diffusion tensor and magnetization transfer MRI measurements of periventricular white matter hyperintensities in old age. *Neurobiology of Aging*, 30(1):125–136.

Bazin, P.-L. and Pham, D. (2004). Topology smoothing for segmentation and surface reconstruction. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, pages 111–118.

Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., and Smith, S. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50(5):1077–1088.

Belaroussi, B., Milles, J., Carme, S., Zhu, Y., and Benoit-Cattin, H. (2006). Intensity non-uniformity correction in MRI: Existing methods and their validation. *Medical Image Analysis*, 10(2):234–246.

Bertalmio, M., Bertozzi, A., and Sapiro, G. (2001). Navier-stokes, fluid dynamics, and image and video inpainting. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 355–362.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36(2):192–326.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 48(3):259–302.

Bezdek, J. and Pal, S. (1991). *Fuzzy Models for Pattern Recognition*. IEEE Press, New Jersey.

Bharath, A. and Ng, J. (2005). A steerable complex wavelet construction and its application to image denoising. *IEEE Transactions on Image Processing*, 14(7):948–959.

Black, M., Sapiro, G., Marimont, D., and Heeger, D. (1998). Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421–432.

Bosc, M., Heitz, F., and Armspach, J. (2003). Statistical atlas-based sub-voxel segmentation of 3D brain MRI. In *Proceedings of the International Conference on Image Processing*, pages 1077–1080.

Brandt, M., Bohan, T., Kramer, L., and Fletcher, J. (1994). Estimation of CSF, white and gray matter volumes in hydrocephalic children using fuzzy clustering of MR images. *Computerized Medical Imaging and Graphics*, 18:25–34.

Bricq, S., Collet, C., and Armspach, J. (2008). Unifying framework for multimodal brain MRI segmentation based on hidden Markov chains. *Computerized Medical Imaging and Graphics*, 29(4):639–652.

Bromiley, P. and Thacker, N. (2008). Improving the accuracy of MR image segmentation through the use of local gradient information. In *Proceedings of Medical Imaging Understanding and Analysis*, pages 84–88.

Brys, G., Hubert, M., and Struyf, A. (2006). Robust measures of tail weight. *Computational Statistics and Data Analysis*, 50(3):733–759.

Buades, A., Coll, B., and Morel, J. (2005). A review of image denoising methods, with a new one. *SIAM Journal of Multiscale Modelling and Simulation*, 4(2):490–530.

Buades, A., Coll, B., and Morel, J. (2006). The staircasing effect in neighborhood filters and its solution. *IEEE Transactions on Image Processing*, 15:1499–1505.

Bullitt, E. and Aylward, S. (2002). Volume rendering of segmented image objects. *IEEE Transactions on Medical Imaging*, 21:998–1002.

Cachia, A., Mangin, J.-F., Riviere, D., Kherif, F., Boddaert, N., Andrade, A., Papadopoulos-Orfanos, D., Poline, J.-B., Bloch, I., Zilbovicius, M., Sonigo, P., Brunelle, F., and Regis, J. (2003). A primal sketch of the cortex mean curvature: A morphogenesis based approach to study variability of the folding patterns. *IEEE Transactions on Medical Imaging*, 22(6):754–765.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698.

Carey, K., Chuang, D., and Hemami, S. (1999). Regularity-preserving image interpolation. *IEEE Transactions on Image Processing*, 8:1293–1297.

Caselles, V., Morel, J., and Sbert, C. (1998). An axiomatic approach to image interpolation. *IEEE Transactions on Image Processing*, 7:376–386.

Catte, F., Lions, P., Morel, J., and Coll, T. (1992). Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193.

Chen, F., Tanaka, K., and Horiguchi, T. (2005). Image segmentation based on Bethe approximation for Gaussian mixture model. *Interdisciplinary Information Sciences*, 11(1):17–29.

Chiverton, J. and Wells, K. (2008). Adaptive partial volume classification of MRI data. *Physics in Medicine and Biology*, 53:5577–5594.

Choi, H., Haynor, D., and Kim, Y. (1991). Partial volume tissue classification of multichannel magnetic resonance images – a mixel model. *IEEE Transactions on Medical Imaging*, 10(3):395–407.

Christensen, S., Wu, O., Hjort, N., Mouridsen, K., Gottrup, C., Fiehler, J., Rother, J., and Ostergaard, L. (2004). Local arterial input functions based on vascular territories. In *Proceedings of the 12th annual meeting of the International Society for Magnetic Resonance in Medicine*, page 363.

Cocosco, C., Kollokian, V., Kwan, R.-S., and Evans, A. (1997). BrainWeb: online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4):S425.

Collins, D., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., and Evans, A. (1998). Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17(3):463–468.

Comaniciu, D. and Meer, P. (2002). Mean-shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.

Courant, R., Friedrichs, K., and Lewy, H. (1928). On the partial difference equations of mathematical physics. *Mathematische Annalen*, 100:32–74.

Criscione, J., Humphrey, J., Douglas, A., and Hunter, W. (2000). An invariant basis for natural strain which yields orthogonal stress response terms in isotropic hyperelasticity. *Journal of the Mechanics and Physics of Solids*, 48(12):2445–2465.

Crossman, A. and Neary, D. (2000). *Neuroanatomy: An Illustrated Colour Text*. Churchill Livingstone.

Crum, W., Hill, D., and Hawkes, D. (2003). Information theoretic similarity measures in non-rigid registration. In *International Conference on Information Processing in Medical Imaging 2003*, pages 378–387.

Dale, A., Fischl, B., and Sereno, M. (1999). Cortical surface-based analysis I: Segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194.

Dalton, C., Brex, P., Jenkins, R., Fox, N., Miszkiel, K., Crum, W., O'Riordan, J., Plant, G., Thompson, A., and Miller, D. (2002). Progressive ventricular enlargement in patients with clinically isolated syndromes is associated with the early development of multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, 73(2):141–147.

Dawant, B., Zijdenbos, A., and Margolin, R. (1993). Correction of intensity variations in MR images for computer-aided tissue classification. *IEEE Transactions on Medical Imaging*, 12:770–781.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38.

Descoteaux, M., Collins, D., and Siddiqi, K. (2004). Geometric flows for segmenting vasculature in MRI: theory and validation. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, pages 500–507.

Descoteaux, M., Collins, D., and Siddiqi, K. (2008). A geometric flow for segmenting vasculature in proton-density weighted MRI. *Medical Image Analysis*, 12:497–513.

Doornik, J. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(s1):927–939.

Dugas-Phocion, G., Ballester, M. G., Malandain, G., Lebrun, C., and Ayache, N. (2004). Improved EM-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 26–33.

Ennis, D. and Kindlmann, G. (2006). Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magnetic Resonance in Medicine*, 55(1):136–146.

Evans, A., Collins, D., Neelin, P., MacDonald, D., Kamber, M., and Marrett, T. (1994). Three-dimensional correlative imaging: applications in human brain mapping. In *Functional Neuroimaging: Technical Foundations*, pages 145–162.

Falcao, A., Udupa, J., Samarasekera, S., and Sharma, S. (1998). User-steered image segmentation paradigms: Livewire and livelane. *Graphical Models and Image Processing*, 60:233–260.

Fiebach, J., Schellinger, P., Geletneky, K., Wilde, P., Meyer, M., Hacke, W., and Sartor, K. (2004). MRI in acute subarachnoid haemorrhage; findings with a standardised stroke protocol. *Neuroradiology*, 46(1):44–48.

Fischl, B. and Dale, A. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences USA*, 97(20):11050–11055.

Fischl, B., Salat, D., van der Kouwe, A., Makris, N., Segonne, F., Quinn, B., and Dale, A. (2004). Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23:S69–S84.

Fischl, B., Sereno, M., and Dale, A. (1999). Cortical surface-based analysis II: Inflation, flattening, and surface-based coordinate system. *NeuroImage*, 9(2):195–207.

Frangi, A., Niessen, W., Vincken, K., and Viergever, M. (1998). Multiscale vessel enhancement filtering. In *Proceedings of Medical Image Computing and Computer-Assisted Interventation*, pages 130–137.

Freeman, W. and Adelson, E. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906.

Fridman, Y., Pizer, S., Aylward, S., and Bullitt, E. (2003). Segmenting 3-D branching tubular structures using cores. In *Proceedings of International Conference on Information Processing in Medical Imaging*, pages 570–577.

Fritsch, D., Eberly, D., Pizer, S., and McAuliffe, M. (1995). Stimulated cores and their applications in medical imaging. In *Proceedings of International Conference on Information Processing in Medical Imaging,*, pages 365–368.

Fu, S., Ruan, Q., Wang, W., and Chen, J. (2006). Region-based fuzzy shock filter with anisotropic diffusion for adaptive image enhancement. In *Intelligent Computing in Signal Processing and Pattern Recognition*, pages 1036–1041.

Gee, J. and Bajcsy, R. (1998). Elastic matching: continuum-mechanical and probabilistic analysis. In *Brain warping*, pages 183–198.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

Gerig, G., Kubler, O., Kikinis, R., and Jolesz, F. (1992). Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232.

Germond, L., Dojat, M., Taylor, C., and Garbay, C. (1999). A multi-agent system for MRI brain segmentation. *Lecture Notes in Computer Science*, 1620:423–432.

Gilboa, G., Sochen, N., and Zeevi, Y. (2002). Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing*, 11(7):689–703.

Gilboa, G., Sochen, N., and Zeevi, Y. (2004). Image enhancement and denoising by complex diffusion processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1020–1036.

Good, C., Johnsrude, I., Ashburner, J., Henson, R., Friston, K., and Frackowiak, R. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, 14(1):21–36.

Goualher, G. L., Procyk, E., Collins, D., Venugopal, R., Barillot, C., and Evans, A. (1999). Automated extraction and variability analysis of sulcal neuroanatomy. *IEEE Transactions on Medical Imaging*, 18(3):206–217.

Gray, H. (1918). *Anatomy of the Human Body*. Philadelphia: Lea & Febiger.

Greenspan, H. (2009). Super-resolution in medical imaging. *The Computer Journal*, 52(1):43–63.

Greenspan, H., Oz, G., Kiryati, N., and Peled, S. (2002). MRI inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging*, 20(5):437–446.

Grimson, W., Ettinger, G., White, S., Gleason, P., Perez, L., Wells, W., and Kikinis, R. (1996). An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. *IEEE Transactions on Medical Imaging*, 15(2):129–140.

Gudbjartsson, H. and Patz, S. (1995). The rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–914.

Haglund, L. (1992). *Adaptive Multidimensional Filtering*. PhD thesis, Linkoping University, Sweden.

Heath, M., Sarkar, S., Sanocki, T., and Bowyer, K. (1997). A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1338–1359.

Herbert, M., Ziegler, D., Deutsch, C., O'Brien, L., Kennedy, D., Filipek, P., Bakardjiev, A., Hodgson, J., Takeoka, M., Makris, N., and Caviness, V. (2005). Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. *Brain*, 128(1):213–226.

Hergan, K., Schaefer, P., Sorensen, A., Gonzalez, R., and Huisman, T. (2002). Diffusion-weighted MRI in diffuse axonal injury of the brain. *European Radiology*, 12(10):2536–2541.

Hua, L. and Yezzi, A. (2007). Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines. *IEEE Transactions on Medical Imaging*, 26(9):1213–1223.

Huang, A., Abugharbieh, R., Tam, R., and Traboulsee, A. (2006). Automatic MRI brain tissue segmentation using a hybrid statistical and geometric model. In *Proceedings of IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pages 394–397.

Jbabdi, S., Woolrich, M., Andersson, J., and Behrens, T. (2007). A Bayesian framework for global tractography. *NeuroImage*, 37:116–129.

Jenkinson, M. and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156.

Jiji, C. and Chaudhuri, S. (2006). Single-frame image super-resolution through contourlet learning. *EURASIP Journal on Applied Signal Processing*, pages 1–11.

Joshi, N. and Brady, M. (2005). Partial volume segmentation of MR images using non-parametric mixture model. In *Proceedings of Medical Image Understanding and Analysis*.

Kakarala, R. and Hero, A. (1992). On achievable accuracy in edge localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7):777–781.

Kirbas, C. and Quek, F. (2003). Vessel extraction techniques and algorithms: a survey. In *Proceedings of Third IEEE Symposium on Bioinformatics and Bioengineering*, pages 238–245.

Knutsson, H., Wilson, R., and Granlund, G. (1983). Anisotropic nonstationary image estimation and its applications: Part I – restoration of noisy images. *IEEE Transactions on Communications*, 31(3):388–397.

Koenderink, J. and van Doorn, A. (1993). Local features of smooth shapes: Ridges and courses. *Geometric Methods in Computer Vision II*, 2031:2–13.

Kovesi, P. (1999). Image features from phase congruency. *Videre: Journal of Computer Vision Research*, 1(3):1–26.

Krissian, K., Malandain, G., Ayache, N., Vaillant, R., and Trousset, Y. (2000). Model-based detection of tubular structures in 3D images. *Computer Vision and Image Understanding*, 80(2):130–171.

Landis, J. and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174.

Lazar, M. and Alexander, A. (2003). An error analysis of white matter tractography methods: synthetic diffusion tensor field simulations. *NeuroImage*, 20:1140–1153.

Lazar, M. and Alexander, A. (2005). Bootstrap white matter tractography (boot-trac). *NeuroImage*, 24:524–532.

Lazar, M., Lee, J., and Alexander, A. (2005). Axial asymmetry of water diffusion in brain white matter. *Magnetic Resonance in Medicine*, 54:860–867.

Lee, J., Yoon, U., Nam, S., Kim, J., Kim, I., and Kim, S. (2003). Evaluation of automated and semi-automated skull-stripping algorithms using similarity index and segmentation error. *Computers in Biology and Medicine*, 33(6):495–507.

Lehmann, T., Gonner, C., and Spitzer, K. (1999). Survey: Interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging*, 18(11):1049–1075.

Lester, H. and Arridge, S. (1999). A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32:129–149.

Li, S., Wang, H., Chan, K., and Petrou, M. (1997). Minimization of MRF energy with relaxation labeling. *Journal of Mathematical Imaging and Vision*, 7(2):149–161.

Liang, Z., Li, X., Eremina, D., and Li, L. (2003). An EM framework for segmentation of tissue mixtures from medical images. In *Proceedings of International Conference of IEEE Engineering in Medicine and Biology*, pages 682–685.

Liang, Z. and Wang, S. (2009). An EM approach to MAP solution of segmenting tissue mixtures: A numerical analysis. *IEEE Transactions on Medical Imaging*, 28(2):297–310.

Liang, Z., Wang, S., Lu, H., and Wang, J. (2007). Model parameter estimation and tissue mixture segmentation by a MAP-EM algorithm. In *IEEE Nuclear Science Symposium Conference Record*, pages 3126–3132.

Likar, B., Viergever, M., and Pernus, F. (2001). Retrospective correction of MR intensity inhomogeneity by information minimization. *IEEE Transactions on Medical Imaging*, 20:1398–1410.

Lindeberg, T. (1993). *Scale-space theory in computer vision*. Kluwer Academic, Hingham.

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):77–116.

Linguraru, M., Ballester, M., and Ayache, N. (2003). A multiscale feature detector for morphological analysis of the brain. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 738–745.

Lopez, A., Lumbreras, F., Serrat, J., and Villanueva, J. (1999). Evaluation of methods for ridge and valley detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):327–335.

Lorigo, L., Faugeras, O., Grimson, W., Keriven, R., Kikinis, R., Nabavi, A., and Westin, C.-F. (2001). CURVES: Curve evolution for vessel segmentation. *Medical Image Analysis*, 5:195–206.

Lorigo, L., Faugeras, O., Grimson, W., Keriven, R., Kikinis, R., Westin, C., and Nabavi, A. (2000). Codimension-two geodesic active contours for the segmentation of tubular structures. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1444–1451.

Ma, J. (2005). Multislice and multicoil phase-sensitive inversion-recovery imaging. *Magnetic Resonance in Medicine*, 53(4):904–910.

Mallat, S. and Hwang, W. (1992). Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38(2):617–643.

Manjon, J., Carbonell-Caballero, J., Lull, J., Garcia-Marti, G., Marti-Bonmati, L., and Robles, M. (2008). MRI denoising using non-local means. *Medical Image Analysis*, 12(4):514–523.

Manjon, J., Lull, J., Carbonell-Caballero, J., Garcia-Marti, G., Marti-Bonmati, L., and Robles, M. (2007). A nonparametric MRI inhomogeneity correction method. *Medical Image Analysis*, 11:336–345.

Manniesing, R. and Niessen, W. (2005). Multiscale vessel enhancing diffusion in CT angiography noise filtering. In *Proceedings of International Conference on Information Processing in Medical Imaging*, pages 138–149.

Manniesing, R., Viergever, M., and Niessen, W. (2006). Vessel enhancing diffusion: a scale space representation of vessel structures. *Medical Image Analysis*, 10(6):815–825.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, 207(1167):187–217.

Mayer, A. and Greenspan, H. (2009). An adaptive mean-shift framework for MRI brain segmentation. *IEEE Transactions on Medical Imaging*, 28(8):1238–1250.

McRobbie, D., Moore, E., Graves, M., and Prince, M. (2007). *MRI from Picture to Proton*. Cambridge University Press.

Mori, S., Crain, B., Chacko, V., and van Zijl, P. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2):265–269.

Morita, T. (1979). Variational principle for the distribution function of the effective field for the random Ising model in the Bethe approximation. *Physica A*, 98:566–572.

Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685.

Nguyen, N. and Milanfar, P. (2000). A wavelet-based interpolation-restoration method for super-resolution. *Circuits, Systems, and Signal Processing*, 19:321–338.

Nyul, L. and Udupa, J. (2002). A protocol-independent brain MRI segmentation method. In *Medical Imaging: Image Processing*, pages 1588–1599.

Ono, M., Kubik, S., and Abernathey, C. (1990). *Atlas of Cerebral Sulci*. Georg Thieme Verlag, Stuttgart.

Openheim, A. and Lim, J. (1981). The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541.

Pearl, J. (1982). Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proceedings of the American Association of Artificial Intelligence National Conference on AI*, pages 133–136.

Penney, G., Schnabel, J., Rueckert, D., Viergever, M., and Niessen, W. (2004). Registration-based interpolation. *IEEE Transactions on Medical Imaging*, 23(7):922–926.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639.

Pham, D. (2001). Spatial models for fuzzy clustering. *Computer Vision and Image Understanding*, 84:285–297.

Pham, D. and Prince, J. (1999). A generalized EM algorithm for robust segmentation of magnetic resonance images. In *Proceedings of The 33rd Annual Conference on Information Sciences and Systems*, pages 558–563.

Pham, D. and Prince, J. (2004). Unsupervised partial volume estimation in single-channel image data. *IEEE Transactions on Medical Imaging*, 23(4):447–458.

Pham, D., Xu, C., and Prince, J. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1):315–337.

Pohl, K., Fisher, J., Grimson, W., and Wells, W. (2005). An expectation-maximisation approach for integrated registration, segmentation, and intensity correction. *Artificial Intelligence Series Publications*, pages 1–13.

Pollak, I. (2002). Segmentation and restoration via non-linear multiscale filtering. *IEEE Signal Processing Magazine*, pages 26–36.

Powell, M. (1996). A review of methods for multivariable interpolation at scattered data points. In *The State of the Art in Numerical Analysis*, pages 283–309.

Prastawa, M., Bullitt, E., Ho, S., and Gerig, G. (2004). A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis*, 8(3):275–283.

Prastawa, M., Bullitt, E., Moon, N., Leemput, K. V., and Gerig, G. (2003). Automatic brain tumor segmentation by subject specific modification of atlas priors. *Academic Radiology*, 10(12):1341–1348.

Pratt, W. (1978). *Digital Image Processing*. Wiley-Interscience, New York.

Renault, C., Desvignes, M., and Revenu, M. (2000). 3D curves tracking and its application to cortical sulci detection. In *International Conference on Image Processing*, pages 491–494.

Roche, A., Malandain, G., Pennec, X., and Ayache, N. (1998). The correlation ratio as a new similarity measure for multimodal image registration. In *Proceedings of MICCAI*, pages 1115–1124.

Rogelj, P. and Kovacic, S. (2001). Similarity measures for non-rigid registration. *Proceedings of the SPIE*, 4322(3):569–578.

Roland, P. and Zilles, K. (1994). Brain atlases – a new research tool. *Trends in Neuroscience*, 17:458–467.

Rosenfeld, A., Hummel, R., and Zucker, S. (1976). Scene labeling by relaxation operations. *IEEE Transactions on Systems, Man and Cybernetics*, 6:420–433.

Royackkers, N., Desvignes, M., Fawal, H., and Revenu, M. (1999). Detection and statistical analysis of human cortical sulci. *NeuroImage*, 10(6):625–641.

Ruan, S., Jaggi, C., Xue, J., Fadili, J., and Bloyet, D. (2000). Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Transactions on Medical Imaging*, 19(12):1179–1187.

Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268.

Salvado, O. (2006). *Characterization of atherosclerosis with magnetic resonance imaging, challenges and validation*. PhD thesis, Case Western Reserve University, Ohio, USA.

Salvado, O., Hillenbrand, C., and Wilson, D. (2006). Partial volume reduction by interpolation with reverse diffusion. *International Journal of Biomedical Imaging*, 1:1–13.

Salvado, O. and Wilson, D. (2006). Thick slice interpolation using reverse anisotropic diffusion to reduce partial volume effect. In *Proceedings of IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pages 1000–1003.

Scherrer, B., Dojat, M., Forbes, F., and Garbay, C. (2007). MRF agent based segmentation: Application to MRI brain scans. In *Artificial Intelligence in Medicine*, pages 13–23.

Schlaug, G., Siewert, B., Benfield, A., Edelman, R., and Warach, S. (1997). Time course of the apparent diffusion coefficient (ADC) abnormality in human stroke. *Neurology*, 49:113–119.

Schnack, H., Pol, H., Baare, W., Viergever, M., and Kahn, R. (2001). Automatic segmentation of the ventricular system from MR images of the human brain. *NeuroImage*, 14(1):95–104.

Shattuck, D., Prasad, G., Mirza, M., Narr, K., and Toga, A. (2009). Online resource for validation of brain segmentation methods. *NeuroImage*, 45(2):431–439.

Shattuck, D., Sandor-Leahy, S., Schaper, K., Rottenberg, D., and Leahy, R. (2001). Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13:856–876.

Shen, S., Sandham, W., Granat, M., and Sterr, A. (2005). MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *IEEE Transactions on Information Technology in Biomedicine*, 9(3):459–467.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

Sim, J. and Wright, C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3):257–268.

Simmons, A., Tofts, P., Barker, G., and Arridge, S. (1994). Sources of intensity nonuniformity in spin echo images at 1.5T. *Magnetic Resonance in Medicine*, 32:121–128.

Sled, J., Zijdenbos, A., and Evans, A. (1998). A nonparametric method for automatic correction of intensity non-uniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17:87–97.

Smith, S. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155.

Sotiropoulos, S., Tench, C., Morgan, P., and Bai, L. (2008). Combining Q-ball imaging and fuzzy connectedness: An approach to distributed brain tractography. In *Proceedings of Medical Image Understanding and Analysis*.

Strikwerda, J. (1989). *Finite Difference Schemes and Partial Differential Equations*. Chapman and Hall.

Studholme, C., Hill, D., and Hawkes, D. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86.

Svensson, B. (2008). *A Multidimensional Filtering Framework with Applications to Local Structure Analysis and Image Enhancement*. PhD thesis, Linkoping University, Sweden.

Svensson, B., Andersson, M., Smedby, O., and Knutsson, H. (2006). Efficient 3-D adaptive filtering for medical image enhancement. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pages 996–999.

Thorstensen, N., Hofer, M., Sapiro, G., and Pottmann, H. (2008). Measuring cortical thickness from volumetric MRI data (unpublished). Unpublished.

Tuch, D. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, 52:1358–1372.

Tuch, D., Reese, T., Wiegell, M., Makris, N., Belliveau, J., and Wedeen, V. (2002). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48:577–582.

Tuch, D., Reese, T., Wiegell, M., and Wedeen, V. (2003). Diffusion MRI of complex neural architecture. *Neuron*, 40:885–895.

Tyrell, J., di Tomaso, E., Fuja, D., Tong, R., Kozak, K., Jain, R., and Roysam, B. (2007). Robust 3-D modeling of vasculature imagery using superellipsoids. *IEEE Transactions on Medical Imaging*, 26(2):223–237.

Udupa, J., Leblanc, V., Zhuge, Y., Imielinska, C., Schmidt, H., Currie, L., Hirsch, B., and Woodburn, J. (2006). A framework for evaluating image segmentation algorithms. *Computerized Medical Imaging and Graphics*, 30(2):75–87.

Udupa, J. and Saha, P. (2003). Fuzzy connectedness and image segmentation. *Proceedings of the IEEE*, 91(10):1649–1669.

van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999). Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908.

van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (2003). A unifying framework for partial volume segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 22(1):105–119.

van Osch, M., Vonken, E., Bakker, C., and Viergever, M. (2001). Correcting partial volume artifacts of the arterial input function in quantitative cerebral perfusion MRI. *Magnetic Resonance in Medicine*, 45:477–485.

Vasilevsky, A. and Siddiqi, K. (2002). Flux maximizing geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1565–1578.

Warfield, S., Zou, K., and Wells, W. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921.

Weickert, J. (1996). Nonlinear diffusion scale-spaces: From the continuous to the discrete setting. In *Proceedings of the 12th International Conference on Analysis and Optimization of Systems Images, Wavelets and PDEs*, pages 111–118.

Weickert, J. (1997). Recursive separable schemes for nonlinear diffusion filters. *Lecture Notes in Computer Science*, 1252:260–271.

Weickert, J. (1998). Fast segmentation methods based on partial differential equations and the watershed transform. In *Proceedings of the DAGM Symposium*, pages 93–100.

Wells, W., Grimson, W., Kikinis, R., and Jolesz, F. (1996). Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442.

Westin, C.-F. and Knutsson, H. (2003). Tensor field regularization using normalized convolution. In *Proceedings of the International Conference on Computer Aided Systems Theory*, pages 564–572.

Wink, O., Niessen, W., and Viergever, M. (2004). Multiscale vessel tracking. *IEEE Transactions on Medical Imaging*, 23(1):130–133.

Withers, J., Bastin, M., and Storkey, A. (2008). Multi-scale segmentation of dual-channel MRI using volume resolution enhancement and tubular structure detection. In *Medical Image Understanding and Analysis*.

Witkin, A. (1983). Scale-space filtering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1019–1021.

Worth, A., Makris, N., Patti, M., Goodman, J., Hoge, E., Caviness, V., and Kennedy, D. (1998). Precise segmentation of the lateral ventricles and caudate nucleus in MR brain images using anatomically driven histograms. *IEEE Transactions on Medical Imaging*, 17(2):303–310.

Wyatt, P. and Noble, J. (2003). MAP MRF joint segmentation and registration. *Medical Image Analysis*, 7(4):539–552.

Yi, J. and Ra, J. (2003). A locally adaptive region growing algorithm for vascular segmentation. *International Journal of Imaging Systems and Technology*, 13(4):208–214.

You, X. and Crebbin, G. (1995). A robust adaptive estimator for filtering noise in images. *IEEE Transactions on Image Processing*, 4(5):693–699.

Yushkevich, P., Piven, J., Hazlett, H., Smith, R., Ho, S., Gee, J., and Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3):1116–1128.

Zaini, M., Schaper, K., Anderson, J., Shattuck, D., Liow, J.-S., Sled, J., and Rottenberg, D. (1999). Method for evaluating the performance of nonuniformity correction algorithms. *NeuroImage*, 9:S99.

Zhang, F. and Hancock, E. (2006). Tensor MRI regularization via graph diffusion. In *Proceedings of the British Machine Vision Conference*, pages 589–598.

Zhang, J. (1992). The mean-field theory in em procedures for Markov random fields. *IEEE Transactions on Signal Processing*, 40(10):2570–2583.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57.

Zheng, S., Tu, Z., Yuille, A., Reiss, A., Dutton, R., Lee, A., Galaburda, A., Thompson, P., Dinov, I., and Toga, A. (2006). A learning based algorithm for automatic extraction of the cortical sulci. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 695–703.

Zhong, J. and Sun, H. (2008). Wavelet-based multiscale anisotropic diffusion with adaptive statistical analysis for image restoration. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 55(9):2716–2725.

Zilles, K., Schleicher, A., Langemann, C., Amunts, K., Morosan, P., Palomero-Gallagher, N., Schormann, T., Mohlberg, H., Burgel, U., Steinmetz, H., Schlaug, G., and Roland, P. (1997). Quantitative analysis of sulci in the human cerebral cortex: Development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. *Human Brain Mapping*, 5(4):218–221.

Zitova, B. and Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing*, 21:977–1000.

Zou, K., Warfield, S., Bharatha, A., Tempany, C., Kaus, M., Haker, S., Wells, W., Jolesz, F., and Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index: scientific reports. *Academic Radiology*, 11(2):178–189.